TABLE 52. Mutagenic Cassette: N, A, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	0
		ALANINE	0		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0	POLAR NONIONIZABLE (POL)	6
		CYSTEINE	0		
AAT	YES	ASPARAGINE	2		
AAC	YES				
CAA	YES	GLUTAMINE	2		
CAG	YES				
TAT	YES	TYROSINE	2		
TAC	YES				
		THREONINE	0		
GAT	YES	ASPARTIC ACID	2	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	4
GAC	YES				
GAA	YES	GLUTAMIC ACID	2		
GAG	YES				
AAA	YES	LYSINE	2	IONIZABLE: BASIC POSITIVE CHARGE (POS)	4
AAG	YES				
		ARGININE	0		
CAT	YES	HISTIDINE	2		
CAC	YES				
TAA	YES	STOP CODON	2	STOP SIGNAL (STP)	2
TAG	YES				
TOTAL	16	7 Amino Acids Are Represented		NPL: POL: NEG: POS: STP = 0: 6: 4: 4: 2	

TABLE 53. Mutagenic Cassette: N, C, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	8
GCT	YES	ALANINE	4		
GCC	YES				
GCA	YES				
GCG	YES				
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
CCT	YES	PROLINE	4		
CCC	YES				
CCA	YES				
CCG	YES				
TCT	YES	SERINE	4	POLAR NONIONIZABLE (POL)	8
TCC	YES				
TCA	YES				
TCG	YES				
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
ACT	YES	THREONINE	4		
ACC	YES				
ACA	YES				
ACG	YES				
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		ARGININE	0		
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	16	4 Amino Acids Are Represented		NPL: POL: NEG: POS: STP = 8: 8: 0: 0: 0	

TABLE 54. Mutagenic Cassette: N, G, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
GGT	YES	GLYCINE	4	NONPOLAR (NPL)	5
GGC	YES				
GGA	YES				
GGG	YES				
		ALANINE	0		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
TGG	YES	TRYPTOPHAN	1		
		PROLINE	0		
AGT	YES	SERINE	2	POLAR NONIONIZABLE (POL)	4
AGC	YES				
TGT	YES	CYSTEINE	2		
TGC	YES				
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	6
CGT	YES	ARGININE	6		
CGC	YES				
CGA	YES				
CGG	YES				
AGA	YES				
AGG	YES				
		HISTIDINE	0		
TGA	YES	STOP CODON	1	STOP SIGNAL (STP)	1
TOTAL	16	5 Amino Acids Are Represented		NPL: POL: NEG: POS: STP = 5: 4: 0: 6: 1	

TABLE 55. Mutagenic Cassette: N, T, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	16
		ALANINE	0		
GTT	YES	VALINE	4		
GTC	YES				
GTA	YES				
GTG	YES				
TTA	YES	LEUCINE	6		
TTG	YES				
CTT	YES				
CTC	YES				
CTA	YES				
CTG	YES				
ATT	YES	ISOLEUCINE	3		
ATC	YES				
ATA	YES				
ATG	YES	METHIONINE	1		
TTT	YES	PHENYLALANINE	2		
TTC	YES				
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0	POLAR NONIONIZABLE (POL)	0
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		ARGININE	0		
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	16	5 Amino Acids Are Represented		NPL: POL: NEG: POS: STP = 16: 0: 0: 0: 0	

TABLE 56. Mutagenic Cassette: N, A/C, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	8
GCT	YES	ALANINE	4		
GCC	YES				
GCA	YES				
GCG	YES				
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
CCT	YES	PROLINE	4		
CCC	YES				
CCA	YES				
CCG	YES				
TCT	YES	SERINE	4	POLAR NONIONIZABLE (POL)	14
TCC	YES				
TCA	YES				
TCG	YES				
		CYSTEINE	0		
AAT	YES	ASPARAGINE	2		
AAC	YES				
CAA	YES	GLUTAMINE	2		
CAG	YES				
TAT	YES	TYROSINE	2		
TAC	YES				
ACT	YES	THREONINE	4		
ACC	YES				
ACA	YES				
ACG	YES				
GAT	YES	ASPARTIC ACID	2	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	4
GAC	YES				
GAA	YES	GLUTAMIC ACID	2		
GAG	YES				
AAA	YES	LYSINE	2	IONIZABLE: BASIC POSITIVE CHARGE (POS)	4
AAG	YES				
		ARGININE	0		
CAT	YES	HISTIDINE	2		
CAC	YES				
TAA	YES	STOP CODON	2	STOP SIGNAL (STP)	2
TAG	YES				
TOTAL	32	11 Amino Acids Are Represented		NPL: 8: POL: 14: NEG: 4: POS: 4: STP = 2	

TABLE 57. Mutagenic Cassette: N, A/G, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
GGT	YES	GLYCINE	4	NONPOLAR (NPL)	5
GGC	YES				
GGA	YES				
GGG	YES				
		ALANINE	0		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
TGG	YES	TRYPTOPHAN	1		
		PROLINE	0		
AGT	YES	SERINE	2	POLAR NONIONIZABLE (POL)	10
AGC	YES				
TGT	YES	CYSTEINE	2		
TGC	YES				
AAT	YES	ASPARAGINE	2		
AAC	YES				
CAA	YES	GLUTAMINE	2		
CAG	YES				
TAT	YES	TYROSINE	2		
TAC	YES				
		THREONINE	0		
GAT	YES	ASPARTIC ACID	2	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	4
GAC	YES				
GAA	YES	GLUTAMIC ACID	2		
GAG	YES				
AAA	YES	LYSINE	2	IONIZABLE: BASIC POSITIVE CHARGE (POS)	10
AAG	YES				
CGT	YES	ARGININE	6		
CGC	YES				
CGA	YES				
CGG	YES				
AGA	YES				
AGG	YES				
CAT	YES	HISTIDINE	2		
CAC	YES				
TAA	YES	STOP CODON	3	STOP SIGNAL (STP)	3
TAG	YES				
TGA	YES				
TOTAL	32	12 Amino Acids Are Represented		NPL: 5: POL: 10: NEG: 4: POS: 10: STP = 3	

TABLE 58. Mutagenic Cassette: N, A/T, N

CODON	Represented	CATEGORY	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	16
		ALANINE	0		
GTT	YES	VALINE	4		
GTC	YES				
GTA	YES				
GTG	YES				
TTA	YES	LEUCINE	6		
TTG	YES				
CTT	YES				
CTC	YES				
CTA	YES				
CTG	YES				
ATT	YES	ISOLEUCINE	3		
ATC	YES				
ATA	YES				
ATG	YES	METHIONINE	1		
TTT	YES	PHENYLALANINE	2		
TTC	YES				
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0	POLAR NONIONIZABLE (POL)	6
		CYSTEINE	0		
AAT	YES	ASPARAGINE	2		
AAC	YES				
CAA	YES	GLUTAMINE	2		
CAG	YES				
TAT	YES	TYROSINE	2		
TAC	YES				
		THREONINE	0		
GAT	YES	ASPARTIC ACID	2	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	4
GAC	YES				
GAA	YES	GLUTAMIC ACID	2		
GAG	YES				
AAA	YES	LYSINE	2	IONIZABLE: BASIC POSITIVE CHARGE (POS)	4
AAG	YES				
		ARGININE	0		
CAT	YES	HISTIDINE	2		
CAC	YES				
TAA	YES	STOP CODON	2	STOP SIGNAL (STP)	2
TAG	YES				
TOTAL	32	12 Amino Acids Are Represented		NPL: POL: NEG: POS: STP = 16: 6: 4: 4: 2	

TABLE 59. Mutagenic Cassette: N, C/G, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
GGT	YES	GLYCINE	4	NONPOLAR (NPL)	13
GGC	YES				
GGA	YES				
GGG	YES				
GCT	YES	ALANINE	4		
GCC	YES				
GCA	YES				
GCG	YES				
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
TGG	YES	TRYPTOPHAN	1		
CCT	YES	PROLINE	4		
CCC	YES				
CCA	YES				
CCG	YES				
TCT	YES	SERINE	6	POLAR NONIONIZABLE (POL)	12
TCC	YES				
TCA	YES				
TCG	YES				
AGT	YES				
AGC	YES				
TGT	YES	CYSTEINE	2		
TGC	YES				
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
ACT	YES	THREONINE	4		
ACC	YES				
ACA	YES				
ACG	YES				
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	6
CGT	YES	ARGININE	6		
CGC	YES				
CGA	YES				
CGG	YES				
AGA	YES				
AGG	YES				
		HISTIDINE	0		
TGA	YES	STOP CODON	1	STOP SIGNAL (STP)	1
TOTAL	32	8 Amino Acids Are Represented		NPL: POL: NEG:POS: STP = 13: 12: 0: 6: 1	

TABLE 60. Mutagenic Cassette: N, C/T, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	24
GCT	YES	ALANINE	4		
GCC	YES				
GCA	YES				
GCG	YES				
GTT	YES	VALINE	4		
GTC	YES				
GTA	YES				
GTG	YES				
TTA	YES	LEUCINE	6		
TTG	YES				
CTT	YES				
CTC	YES				
CTA	YES				
CTG	YES				
ATT	YES	ISOLEUCINE	3		
ATC	YES				
ATA	YES				
ATG	YES	METHIONINE	1		
TTT	YES	PHENYLALANINE	2		
TTC	YES				
		TRYPTOPHAN	0		
CCT	YES	PROLINE	4		
CCC	YES				
CCA	YES				
CCG	YES				
TCT	YES	SERINE	4	POLAR NONIONIZABLE (POL)	8
TCC	YES				
TCA	YES				
TCG	YES				
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
ACT	YES	THREONINE	4		
ACC	YES				
ACA	YES				
ACG	YES				
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		ARGININE	0		
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	32	9 Amino Acids Are Represented		NPL: 24: POL: 8: NEG: 0: POS: 0: 0	STP =

TABLE 61. Mutagenic Cassette: N, G/T, N

	CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
	GGT	YES	GLYCINE	4	NONPOLAR (NPL)	21
	GGC	YES				
	GGA	YES				
	GGG	YES				
			ALANINE	0		
	GTT	YES	VALINE	4		
	GTC	YES				
	GTA	YES				
	GTG	YES				
	TTA	YES	LEUCINE	6		
	TTG	YES				
	CTT	YES				
	CTC	YES				
	CTA	YES				
	CTG	YES				
	ATT	YES	ISOLEUCINE	3		
	ATC	YES				
	ATA	YES				
	ATG	YES	METHIONINE	1		
	TTT	YES	PHENYLALANINE	2		
	TTC	YES				
	TGG	YES	TRYPTOPHAN	1		
			PROLINE	0		
	AGT	YES	SERINE	2	POLAR NONIONIZABLE (POL)	4
	AGC	YES				
	TGT	YES	CYSTEINE	2		
	TGC	YES				
			ASPARAGINE	0		
			GLUTAMINE	0		
			TYROSINE	0		
			THREONINE	0		
			ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
			GLUTAMIC ACID	0		
			LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	6
	CGT	YES	ARGININE	6		
	CGC	YES				
	CGA	YES				
	CGG	YES				
	AGA	YES				
	AGG	YES				
			HISTIDINE	0		
	TGA	YES	STOP CODON	1	STOP SIGNAL (STP)	1
TOTAL		32	10 Amino Acids Are Represented		NPL: POL: NEG:POS: STP = 21: 4: 0: 6: 1	

TABLE 62. Mutagenic Cassette: N, A/C/G, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
GGT	YES	GLYCINE	4	NONPOLAR (NPL)	13
GGC	YES				
GGA	YES				
GGG	YES				
GCT	YES	ALANINE	4		
GCC	YES				
GCA	YES				
GCG	YES				
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
TGG	YES	TRYPTOPHAN	1		
CCT	YES	PROLINE	4		
CCC	YES				
CCA	YES				
CCG	YES				
TCT	YES	SERINE	6	POLAR NONIONIZABLE (POL)	18
TCC	YES				
TCA	YES				
TCG	YES				
AGT	YES				
AGC	YES				
TGT	YES	CYSTEINE	2		
TGC	YES				
AAT	YES	ASPARAGINE	2		
AAC	YES				
CAA	YES	GLUTAMINE	2		
CAG	YES				
TAT	YES	TYROSINE	2		
TAC	YES				
ACT	YES	THREONINE	4		
ACC	YES				
ACA	YES				
ACG	YES				
GAT	YES	ASPARTIC ACID	2	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	4
GAC	YES				
GAA	YES	GLUTAMIC ACID	2		
GAG	YES				
AAA	YES	LYSINE	2	IONIZABLE: BASIC POSITIVE CHARGE (POS)	10
AAG	YES				
CGT	YES	ARGININE	6		
CGC	YES				
CGA	YES				
CGG	YES				
AGA	YES				
AGG	YES				
CAT	YES	HISTIDINE	2		
CAC	YES				
TAA	YES	STOP CODON	3	STOP SIGNAL (STP)	3
TAG	YES				
TGA	YES				
TOTAL	48	15 Amino Acids Are Represented		NPL: POL: NEG:POS: STP = 13: 18: 4: 10: 3	

TABLE 63. Mutagenic Cassette: N, A/C/T, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	24
GCT	YES	ALANINE	4		
GCC	YES				
GCA	YES				
GCG	YES				
GTT	YES	VALINE	4		
GTC	YES				
GTA	YES				
GTG	YES				
TTA	YES	LEUCINE	6		
TTG	YES				
CTT	YES				
CTC	YES				
CTA	YES				
CTG	YES				
ATT	YES	ISOLEUCINE	3		
ATC	YES				
ATA	YES				
ATG	YES	METHIONINE	1		
TTT	YES	PHENYLALANINE	2		
TTC	YES				
		TRYPTOPHAN	0		
CCT	YES	PROLINE	4		
CCC	YES				
CCA	YES				
CCG	YES				
TCT	YES	SERINE	4	POLAR NONIONIZABLE (POL)	14
TCC	YES				
TCA	YES				
TCG	YES				
		CYSTEINE	0		
AAT	YES	ASPARAGINE	2		
AAC	YES				
CAA	YES	GLUTAMINE	2		
CAG	YES				
TAT	YES	TYROSINE	2		
TAC	YES				
ACT	YES	THREONINE	4		
ACC	YES				
ACA	YES				
ACG	YES				
GAT	YES	ASPARTIC ACID	2	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	4
GAC	YES				
GAA	YES	GLUTAMIC ACID	2		
GAG	YES				
AAA	YES	LYSINE	2	IONIZABLE: BASIC POSITIVE CHARGE (POS)	4
AAG	YES				
		ARGININE	0		
CAT	YES	HISTIDINE	2		
CAC	YES				
TAA	YES	STOP CODON	2	STOP SIGNAL (STP)	2
TAG	YES				
TOTAL	48	16 Amino Acids Are Represented		NPL: POL: NEG:POS: STP = 24: 14: 4: 4: 2	

TABLE 64. Mutagenic Cassette: N, A/G/T, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
GGT	YES	GLYCINE	4	NONPOLAR (NPL)	21
GGC	YES				
GGA	YES				
GGG	YES				
		ALANINE	0		
GTT	YES	VALINE	4		
GTC	YES				
GTA	YES				
GTG	YES				
TTA	YES	LEUCINE	6		
TTG	YES				
CTT	YES				
CTC	YES				
CTA	YES				
CTG	YES				
ATT	YES	ISOLEUCINE	3		
ATC	YES				
ATA	YES				
ATG	YES	METHIONINE	1		
TTT	YES	PHENYLALANINE	2		
TTC	YES				
TGG	YES	TRYPTOPHAN	1		
		PROLINE	0		
AGT	YES	SERINE	2	POLAR NONIONIZABLE (POL)	10
AGC	YES				
TGT	YES	CYSTEINE	2		
TGC	YES				
AAT	YES	ASPARAGINE	2		
AAC	YES				
CAA	YES	GLUTAMINE	2		
CAG	YES				
TAT	YES	TYROSINE	2		
TAC	YES				
		THREONINE	0		
GAT	YES	ASPARTIC ACID	2	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	4
GAC	YES				
GAA	YES	GLUTAMIC ACID	2		
GAG	YES				
AAA	YES	LYSINE	2	IONIZABLE: BASIC POSITIVE CHARGE (POS)	10
AAG	YES				
CGT	YES	ARGININE	6		
CGC	YES				
CGA	YES				
CGG	YES				
AGA	YES				
AGG	YES				
CAT	YES	HISTIDINE	2		
CAC	YES				
TAA	YES	STOP CODON	3	STOP SIGNAL (STP)	3
TAG	YES				
TGA	YES				
TOTAL	48	17 Amino Acids Are Represented		NPL: POL: NEG:POS: STP = 21: 10: 4: 10: 3	

TABLE 65. Mutagenic Cassette: N, C/G/T, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
GGT	YES	GLYCINE	4	NONPOLAR (NPL)	29
GGC	YES				
GGA	YES				
GGG	YES				
GCT	YES	ALANINE	4		
GCC	YES				
GCA	YES				
GCG	YES				
GTT	YES	VALINE	4		
GTC	YES				
GTA	YES				
GTG	YES				
TTA	YES	LEUCINE	6		
TTG	YES				
CTT	YES				
CTC	YES				
CTA	YES				
CTG	YES				
ATT	YES	ISOLEUCINE	3		
ATC	YES				
ATA	YES				
ATG	YES	METHIONINE	1		
TTT	YES	PHENYLALANINE	2		
TTC	YES				
TGG	YES	TRYPTOPHAN	1		
CCT	YES	PROLINE	4		
CCC	YES				
CCA	YES				
CCG	YES				
TCT	YES	SERINE	6	POLAR NONIONIZABLE (POL)	12
TCC	YES				
TCA	YES				
TCG	YES				
AGT	YES				
AGC	YES				
TGT	YES	CYSTEINE	2		
TGC	YES				
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
ACT	YES	THREONINE	4		
ACC	YES				
ACA	YES				
ACG	YES				
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	6
CGT	YES	ARGININE	6		
CGC	YES				
CGA	YES				
CGG	YES				
AGA	YES				
AGG	YES				
		HISTIDINE	0		
TGA	YES	STOP CODON	1	STOP SIGNAL (STP)	1
TOTAL	48	13 Amino Acids Are Represented		NPL: POL: NEG:POS: STP = 29: 12: 0: 6: 1	

TABLE 66. Mutagenic Cassette: C, C, N

	CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
			GLYCINE	0	NONPOLAR (NPL)	4
			ALANINE	0		
			VALINE	0		
			LEUCINE	0		
			ISOLEUCINE	0		
			METHIONINE	0		
			PHENYLALANINE	0		
			TRYPTOPHAN	0		
	CCT	YES	PROLINE	4		
	CCC	YES				
	CCA	YES				
	CCG	YES				
			SERINE	0	POLAR NONIONIZABLE (POL)	0
			CYSTEINE	0		
			ASPARAGINE	0		
			GLUTAMINE	0		
			TYROSINE	0		
			THREONINE	0		
			ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
			GLUTAMIC ACID	0		
			LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
			ARGININE	0		
			HISTIDINE	0		
			STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL		4	1 Amino Acid is Represented		NPL: POL: NEG:POS: STP = 4: 0: 0: 0: 0	

TABLE 67. Mutagenic Cassette: G, G, N

	CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
	GGT	YES	GLYCINE	4	NONPOLAR (NPL)	4
	GGC	YES				
	GGA	YES				
	GGG	YES				
			ALANINE	0		
			VALINE	0		
			LEUCINE	0		
			ISOLEUCINE	0		
			METHIONINE	0		
			PHENYLALANINE	0		
			TRYPTOPHAN	0		
			PROLINE	0		
			SERINE	0	POLAR NONIONIZABLE (POL)	0
			CYSTEINE	0		
			ASPARAGINE	0		
			GLUTAMINE	0		
			TYROSINE	0		
			THREONINE	0		
			ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
			GLUTAMIC ACID	0		
			LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
			ARGININE	0		
			HISTIDINE	0		
			STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	0	4	1 Amino Acid is Represented		NPL: POL: NEG:POS: STP = 4: 0: 0: 0: 0	

TABLE 68. Mutagenic Cassette: G, C, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	4
GCT	YES	ALANINE	4		
GCC	YES				
GCA	YES				
GCG	YES				
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0	POLAR NONIONIZABLE (POL)	0
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		ARGININE	0		
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	4	1 Amino Acid Is Represented		NPL: 4: POL: 0: NEG: 0: POS: 0: STP = 0	

TABLE 69. Mutagenic Cassette: G, T, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	4
		ALANINE	0		
GTT	YES	VALINE	4		
GTC	YES				
GTA	YES				
GTG	YES				
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0	POLAR NONIONIZABLE (POL)	0
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		ARGININE	0		
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	4	1 Amino Acid Is Represented		NPL: 4: POL: 0: NEG: 0: POS: 0: STP = 0	

TABLE 70. Mutagenic Cassette: C, G, N

	CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
			GLYCINE	0	NONPOLAR (NPL)	0
			ALANINE	0		
			VALINE	0		
			LEUCINE	0		
			ISOLEUCINE	0		
			METHIONINE	0		
			PHENYLALANINE	0		
			TRYPTOPHAN	0		
			PROLINE	0		
			SERINE	0	POLAR NONIONIZABLE (POL)	0
			CYSTEINE	0		
			ASPARAGINE	0		
			GLUTAMINE	0		
			TYROSINE	0		
			THREONINE	0		
			ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
			GLUTAMIC ACID	0		
			LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	4
	CGT	YES	ARGININE	4		
	CGC	YES				
	CGA	YES				
	CGG	YES				
			HISTIDINE	0		
			STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL		4	1 Amino Acid Is Represented		NPL: POL: NEG:POS: STP = 0: 0: 0: 4: 0	

TABLE 71. Mutagenic Cassette: C, T, N

	CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
			GLYCINE	0	NONPOLAR (NPL)	4
			ALANINE	0		
			VALINE	0		
	CTT	YES	LEUCINE	4		
	CTC	YES				
	CTA	YES				
	CTG	YES				
			ISOLEUCINE	0		
			METHIONINE	0		
			PHENYLALANINE	0		
			TRYPTOPHAN	0		
			PROLINE	0		
			SERINE	0	POLAR NONIONIZABLE (POL)	0
			CYSTEINE	0		
			ASPARAGINE	0		
			GLUTAMINE	0		
			TYROSINE	0		
			THREONINE	0		
			ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
			GLUTAMIC ACID	0		
			LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
			ARGININE	0		
			HISTIDINE	0		
			STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL		4	1 Amino Acid Is Represented		NPL: POL: NEG:POS: STP = 4: 0: 0: 0: 0	

TABLE 72. Mutagenic Cassette: T, C, N

	CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
			GLYCINE	0	NONPOLAR (NPL)	0
			ALANINE	0		
			VALINE	0		
			LEUCINE	0		
			ISOLEUCINE	0		
			METHIONINE	0		
			PHENYLALANINE	0		
			TRYPTOPHAN	0		
			PROLINE	0		
	TCT	YES	SERINE	4	POLAR NONIONIZABLE (POL)	4
	TCC	YES				
	TCA	YES				
	TCG	YES				
			CYSTEINE	0		
			ASPARAGINE	0		
			GLUTAMINE	0		
			TYROSINE	0		
			THREONINE	0		
			ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
			GLUTAMIC ACID	0		
			LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
			ARGININE	0		
			HISTIDINE	0		
			STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL		4	1 Amino Acid Is Represented		NPL: POL: NEG: POS: STP = 0: 4: 0: 0: 0	

TABLE 73. Mutagenic Cassette: A, C, N

	CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
			GLYCINE	0	NONPOLAR (NPL)	0
			ALANINE	0		
			VALINE	0		
			LEUCINE	0		
			ISOLEUCINE	0		
			METHIONINE	0		
			PHENYLALANINE	0		
			TRYPTOPHAN	0		
			PROLINE	0		
			SERINE	0	POLAR NONIONIZABLE (POL)	4
			CYSTEINE	0		
			ASPARAGINE	0		
			GLUTAMINE	0		
			TYROSINE	0		
	ACT	YES	THREONINE	4		
	ACC	YES				
	ACA	YES				
	ACG	YES				
			ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
			GLUTAMIC ACID	0		
			LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
			ARGININE	0		
			HISTIDINE	0		
			STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL		4	1 Amino Acid Is Represented		NPL: POL: NEG: POS: STP = 0: 4: 0: 0: 0	

TABLE 74. Mutagenic Cassette: G, A, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	0
		ALANINE	0		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0	POLAR NONIONIZABLE (POL)	0
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0		
GAT	YES	ASPARTIC ACID	2	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	4
GAC	YES				
GAA	YES	GLUTAMIC ACID	2		
GAG	YES				
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		ARGININE	0		
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	4	2 Amino Acids Are Represented		NPL: 0: POL: 0: NEG: 4: POS: 0: STP = 0	

TABLE 75. Mutagenic Cassette: A, T, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	4
		ALANINE	0		
		VALINE	0		
		LEUCINE	0		
ATT	YES	ISOLEUCINE	3		
ATC	YES				
ATA	YES				
ATG	YES	METHIONINE	1		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0	POLAR NONIONIZABLE (POL)	0
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		ARGININE	0		
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	4	2 Amino Acids Are Represented		NPL: 4: POL: 0: NEG: 0: POS: 0: STP = 0	

TABLE 76. Mutagenic Cassette: C, A, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	0
		ALANINE	0		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0	POLAR NONIONIZABLE (POL)	2
		CYSTEINE	0		
		ASPARAGINE	0		
CAA	YES	GLUTAMINE	2		
CAG	YES				
		TYROSINE	0		
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	2
		ARGININE	0		
CAT	YES	HISTIDINE	2		
CAC	YES				
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	4	2 Amino Acids Are Represented		NPL: POL: NEG:POS: STP = 0: 2: 0: 2: 0	

TABLE 77. Mutagenic Cassette: T, T, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	4
		ALANINE	0		
		VALINE	0		
TTA	YES	LEUCINE	2		
TTG	YES				
		ISOLEUCINE	0		
		METHIONINE	0		
TTT	YES	PHENYLALANINE	2		
TTC	YES				
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0	POLAR NONIONIZABLE (POL)	0
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		ARGININE	0		
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	4	2 Amino Acids Are Represented		NPL: POL: NEG:POS: STP = 4: 0: 0: 0: 0	

TABLE 78. Mutagenic Cassette: A, A, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	0
		ALANINE	0		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0	POLAR NONIONIZABLE (POL)	2
		CYSTEINE	0		
AAT	YES	ASPARAGINE	2		
AAC	YES				
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
AAA	YES	LYSINE	2	IONIZABLE: BASIC POSITIVE CHARGE (POS)	2
AAG	YES				
		ARGININE	0		
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	4	2 Amino Acids Are Represented		NPL: 0: POL: 2: NEG:0: POS: 2: STP = 0	

TABLE 79. Mutagenic Cassette: T, A, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	0
		ALANINE	0		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0	POLAR NONIONIZABLE (POL)	2
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
TAT	YES	TYROSINE	2		
TAC	YES				
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		ARGININE	0		
		HISTIDINE	0		
TAA	YES	STOP CODON	2	STOP SIGNAL (STP)	2
TAG	YES				
TOTAL	4	1 Amino Acid Is Represented		NPL: 0: POL: 2: NEG:0: POS: 0: STP = 2	

TABLE 80. Mutagenic Cassette: T, G, N

	CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
			GLYCINE	0	NONPOLAR (NPL)	1
			ALANINE	0		
			VALINE	0		
			LEUCINE	0		
			ISOLEUCINE	0		
			METHIONINE	0		
			PHENYLALANINE	0		
	TGG	YES	TRYPTOPHAN	1		
			PROLINE	0		
			SERINE	0	POLAR NONIONIZABLE (POL)	2
	TGT	YES	CYSTEINE	2		
	TGC	YES				
			TYROSINE	0		
			THREONINE	0		
			ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
			GLUTAMIC ACID	0		
			LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
			ARGININE	0		
			HISTIDINE	0		
	TGA	YES	STOP CODON	1	STOP SIGNAL (STP)	1
TOTAL		4	2 Amino Acids Are Represented		NPL: 1: POL: 2: NEG: 0: POS: 0: STP = 1	

TABLE 81. Mutagenic Cassette: A, G, N

	CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
			GLYCINE	0	NONPOLAR (NPL)	0
			ALANINE	0		
			VALINE	0		
			LEUCINE	0		
			ISOLEUCINE	0		
			METHIONINE	0		
			PHENYLALANINE	0		
			TRYPTOPHAN	0		
			PROLINE	0		
	AGT	YES	SERINE	2	POLAR NONIONIZABLE (POL)	2
	AGC	YES				
			CYSTEINE	0		
			ASPARAGINE	0		
			GLUTAMINE	0		
			TYROSINE	0		
			THREONINE	0		
			ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
			GLUTAMIC ACID	0		
			LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	2
	AGA	YES	ARGININE	2		
	AGG	YES				
			HISTIDINE	0		
			STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL		4	2 Amino Acids Are Represented		NPL: 0: POL: 2: NEG: 0: POS: 2: STP = 0	

TABLE 82. Mutagenic Cassette: G/C, G, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
GGT	YES	GLYCINE	4	NONPOLAR (NPL)	4
GGC	YES				
GGA	YES				
GGG	YES				
		ALANINE	0		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0	POLAR NONIONIZABLE (POL)	0
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	4
CGT	YES	ARGININE	4		
CGC	YES				
CGA	YES				
CGG	YES				
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL		8	2 Amino Acids Are Represented	NPL: 4: POL: 0: NEG: 0: POS: 4: STP = 0	

TABLE 83. Mutagenic Cassette: G/C, C, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	8
GCT	YES	ALANINE	4		
GCC	YES				
GCA	YES				
GCG	YES				
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
CCT	YES	PROLINE	4		
CCC	YES				
CCA	YES				
CCG	YES				
		SERINE	0	POLAR NONIONIZABLE (POL)	0
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		ARGININE	0		
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL		8	2 Amino Acids Are Represented	NPL: 8: POL: 0: NEG: 0: POS: 0: STP = 0	

TABLE 84. Mutagenic Cassette: G/C, A, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	0
		ALANINE	0		
		VALINE	0		
		LEUCINE	0		
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0	POLAR NONIONIZABLE (POL)	2
		CYSTEINE	0		
		ASPARAGINE	0		
CAA	YES	GLUTAMINE	2		
CAG	YES				
		TYROSINE	0		
		THREONINE	0		
GAT	YES	ASPARTIC ACID	2	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	4
GAC	YES				
GAA	YES	GLUTAMIC ACID	2		
GAG	YES				
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	2
		ARGININE	0		
CAT	YES	HISTIDINE	2		
CAC	YES				
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	8	4 Amino Acids Are Represented		NPL: 0: POL: 2: NEG: 4: POS: 2: STP = 0	

TABLE 85. Mutagenic Cassette: G/C, T, N

CODON	Represented	AMINO ACID	(Frequency)	CATEGORY	(Frequency)
		GLYCINE	0	NONPOLAR (NPL)	8
		ALANINE	0		
GTT	YES	VALINE	4		
GTC	YES				
GTA	YES				
GTG	YES				
CTT	YES	LEUCINE	4		
CTC	YES				
CTA	YES				
CTG	YES				
		ISOLEUCINE	0		
		METHIONINE	0		
		PHENYLALANINE	0		
		TRYPTOPHAN	0		
		PROLINE	0		
		SERINE	0	POLAR NONIONIZABLE (POL)	0
		CYSTEINE	0		
		ASPARAGINE	0		
		GLUTAMINE	0		
		TYROSINE	0		
		THREONINE	0		
		ASPARTIC ACID	0	IONIZABLE: ACIDIC NEGATIVE CHARGE (NEG)	0
		GLUTAMIC ACID	0		
		LYSINE	0	IONIZABLE: BASIC POSITIVE CHARGE (POS)	0
		ARGININE	0		
		HISTIDINE	0		
		STOP CODON	0	STOP SIGNAL (STP)	0
TOTAL	8	2 Amino Acids Are Represented		NPL: 8: POL: 0: NEG: 0: POS: 0: STP = 0	

2.11.2. CHIMERIZATIONS

2.11.2.1 "SHUFFLING"

Nucleic acid shuffling is a method for *in vitro* or *in vivo* homologous recombination of pools of shorter or smaller polynucleotides to produce a polynucleotide or polynucleotides. Mixtures of related nucleic acid sequences or polynucleotides are subjected to sexual PCR to provide random polynucleotides, and reassembled to yield a library or mixed population of recombinant hybrid nucleic acid molecules or polynucleotides.

In contrast to cassette mutagenesis, only shuffling and error-prone PCR allow one to mutate a pool of sequences blindly (without sequence information other than primers).

The advantage of the mutagenic shuffling of this invention over error-prone PCR alone for repeated selection can best be explained with an example from antibody engineering. Consider DNA shuffling as compared with error-prone PCR (not sexual PCR). The initial library of selected pooled sequences can consist of related sequences of diverse origin (i.e. antibodies from naive mRNA) or can be derived by any type of mutagenesis (including shuffling) of a single antibody gene. A collection of selected complementarity determining regions ("CDRs") is obtained after the first round of affinity selection. In the diagram the thick CDRs confer onto the antibody molecule increased affinity for the antigen. Shuffling allows the free combinatorial association of all of the CDR1s with all of the CDR2s with all of the CDR3s, for example.

This method differs from error-prone PCR, in that it is an inverse chain reaction. In error-prone PCR, the number of polymerase start sites and the number of molecules grows exponentially. However, the sequence of the polymerase start sites and the sequence of the molecules remains essentially the same. In contrast, in nucleic acid reassembly or shuffling of random polynucleotides the number of start sites and the

number (but not size) of the random polynucleotides decreases over time. For polynucleotides derived from whole plasmids the theoretical endpoint is a single, large concatemeric molecule.

Since cross-overs occur at regions of homology, recombination will primarily occur between members of the same sequence family. This discourages combinations of CDRs that are grossly incompatible (e.g., directed against different epitopes of the same antigen). It is contemplated that multiple families of sequences can be shuffled in the same reaction. Further, shuffling generally conserves the relative order, such that, for example, CDR1 will not be found in the position of CDR2.

Rare shufflants will contain a large number of the best (eg. highest affinity) CDRs and these rare shufflants may be selected based on their superior affinity.

CDRs from a pool of 100 different selected antibody sequences can be permutated in up to 1006 different ways. This large number of permutations cannot be represented in a single library of DNA sequences. Accordingly, it is contemplated that multiple cycles of DNA shuffling and selection may be required depending on the length of the sequence and the sequence diversity desired.

Error-prone PCR, in contrast, keeps all the selected CDRs in the same relative sequence, generating a much smaller mutant cloud.

The template polynucleotide which may be used in the methods of this invention may be DNA or RNA. It may be of various lengths depending on the size of the gene or shorter or smaller polynucleotide to be recombined or reassembled. Preferably, the template polynucleotide is from 50 bp to 50 kb. It is contemplated that entire vectors containing the nucleic acid encoding the protein of interest can be used in the methods of this invention, and in fact have been successfully used.

The template polynucleotide may be obtained by amplification using the PCR reaction (**USPN 4,683,202 and USPN 4,683,195**) or other amplification or cloning methods. However, the removal of free primers from the PCR products before subjecting them to pooling of the PCR products and sexual PCR may provide more efficient results. Failure to adequately remove the primers from the original pool before sexual PCR can lead to a low frequency of crossover clones.

The template polynucleotide often should be double-stranded. A double-stranded nucleic acid molecule is recommended to ensure that regions of the resulting single-stranded polynucleotides are complementary to each other and thus can hybridize to form a double-stranded molecule.

It is contemplated that single-stranded or double-stranded nucleic acid polynucleotides having regions of identity to the template polynucleotide and regions of heterology to the template polynucleotide may be added to the template polynucleotide, at this step. It is also contemplated that two different but related polynucleotide templates can be mixed at this step.

The double-stranded polynucleotide template and any added double-or single-stranded polynucleotides are subjected to sexual PCR which includes slowing or halting to provide a mixture of from about 5 bp to 5 kb or more. Preferably the size of the random polynucleotides is from about 10 bp to 1000 bp, more preferably the size of the polynucleotides is from about 20 bp to 500 bp.

Alternatively, it is also contemplated that double-stranded nucleic acid having multiple nicks may be used in the methods of this invention. A nick is a break in one strand of the double-stranded nucleic acid. The distance between such nicks is preferably 5 bp to 5 kb, more preferably between 10 bp to 1000 bp. This can provide areas of self-

priming to produce shorter or smaller polynucleotides to be included with the polynucleotides resulting from random primers, for example.

The concentration of any one specific polynucleotide will not be greater than 1% by weight of the total polynucleotides, more preferably the concentration of any one specific nucleic acid sequence will not be greater than 0.1% by weight of the total nucleic acid.

The number of different specific polynucletides in the mixture will be at least about 100, preferably at least about 500, and more preferably at least about 1000.

At this step single-stranded or double-stranded polynucleotides, either synthetic or natural, may be added to the random double-stranded shorter or smaller polynucleotides in order to increase the heterogeneity of the mixture of polynucleotides.

It is also contemplated that populations of double-stranded randomly broken polynucleotides may be mixed or combined at this step with the polynucleotides from the sexual PCR process and optionally subjected to one or more additional sexual PCR cycles.

Where insertion of mutations into the template polynucleotide is desired, single-stranded or double-stranded polynucleotides having a region of identity to the template polynucleotide and a region of heterology to the template polynucleotide may be added in a 20 fold excess by weight as compared to the total nucleic acid, more preferably the single-stranded polynucleotides may be added in a 10 fold excess by weight as compared to the total nucleic acid.

Where a mixture of different but related template polynucleotides is desired, populations of polynucleotides from each of the templates may be combined at a ratio of

less than about 1:100, more preferably the ratio is less than about 1:40. For example, a backcross of the wild-type polynucleotide with a population of mutated polynucleotide may be desired to eliminate neutral mutations (e.g., mutations yielding an insubstantial alteration in the phenotypic property being selected for). In such an example, the ratio of randomly provided wild-type polynucleotides which may be added to the randomly provided sexual PCR cycle hybrid polynucleotides is approximately 1:1 to about 100:1, and more preferably from 1:1 to 40:1.

The mixed population of random polynucleotides are denatured to form single-stranded polynucleotides and then re-annealed. Only those single-stranded polynucleotides having regions of homology with other single-stranded polynucleotides will re-anneal.

The random polynucleotides may be denatured by heating. One skilled in the art could determine the conditions necessary to completely denature the double-stranded nucleic acid. Preferably the temperature is from 80 °C to 100 °C, more preferably the temperature is from 90 °C to 96 °C. other methods which may be used to denature the polynucleotides include pressure (36) and pH.

The polynucleotides may be re-annealed by cooling. Preferably the temperature is from 20 °C to 75 °C, more preferably the temperature is from 40 °C to 65 °C. If a high frequency of crossovers is needed based on an average of only 4 consecutive bases of homology, recombination can be forced by using a low annealing temperature, although the process becomes more difficult. The degree of renaturation which occurs will depend on the degree of homology between the population of single-stranded polynucleotides.

Renaturation can be accelerated by the addition of polyethylene glycol ("PEG") or salt. The salt concentration is preferably from 0 mM to 200 mM, more preferably the salt

concentration is from 10 mM to 100 mm. The salt may be KCl or NaCl. The concentration of PEG is preferably from 0% to 20%, more preferably from 5% to 10%.

The annealed polynucleotides are next incubated in the presence of a nucleic acid polymerase and dNTP's (i.e. dATP, dCTP, DGTP and dTTP). The nucleic acid polymerase may be the Klenow fragment, the Taq polymerase or any other DNA polymerase known in the art.

The approach to be used for the assembly depends on the minimum degree of homology that should still yield crossovers. If the areas of identity are large, Taq polymerase can be used with an annealing temperature of between 45-65 °C. If the areas of identity are small, Klenow polymerase can be used with an annealing temperature of between 20-30 °C. One skilled in the art could vary the temperature of annealing to increase the number of cross-overs achieved.

The polymerase may be added to the random polynucleotides prior to annealing, simultaneously with annealing or after annealing.

The cycle of denaturation, renaturation and incubation in the presence of polymerase is referred to herein as shuffling or reassembly of the nucleic acid. This cycle is repeated for a desired number of times. Preferably the cycle is repeated from 2 to 50 times, more preferably the sequence is repeated from 10 to 40 times.

The resulting nucleic acid is a larger double-stranded polynucleotide of from about 50 bp to about 100 kb, preferably the larger polynucleotide is from 500 bp to 50 kb.

This larger polynucleotides may contain a number of copies of a polynucleotide having the same size as the template polynucleotide in tandem. This concatemeric polynucleotide is then denatured into single copies of the template polynucleotide. The

result will be a population of polynucleotides of approximately the same size as the template polynucleotide. The population will be a mixed population where single or double-stranded polynucleotides having an area of identity and an area of heterology have been added to the template polynucleotide prior to shuffling. These polynucleotides are then cloned into the appropriate vector and the ligation mixture used to transform bacteria.

It is contemplated that the single polynucleotides may be obtained from the larger concatemeric polynucleotide by amplification of the single polynucleotide prior to cloning by a variety of methods including PCR (**USPN 4,683,195 and USPN 4,683,202**), rather than by digestion of the concatemer.

The vector used for cloning is not critical provided that it will accept a polynucleotide of the desired size. If expression of the particular polynucleotide is desired, the cloning vehicle should further comprise transcription and translation signals next to the site of insertion of the polynucleotide to allow expression of the polynucleotide in the host cell. Preferred vectors include the pUC series and the pBR series of plasmids.

The resulting bacterial population will include a number of recombinant polynucleotides having random mutations. This mixed population may be tested to identify the desired recombinant polynucleotides. The method of selection will depend on the polynucleotide desired.

For example, if a polynucleotide which encodes a protein with increased binding efficiency to a ligand is desired, the proteins expressed by each of the portions of the polynucleotides in the population or library may be tested for their ability to bind to the ligand by methods known in the art (i.e. panning, affinity chromatography). If a polynucleotide which encodes for a protein with increased drug resistance is desired, the

proteins expressed by each of the polynucleotides in the population or library may be tested for their ability to confer drug resistance to the host organism. One skilled in the art, given knowledge of the desired protein, could readily test the population to identify polynucleotides which confer the desired properties onto the protein.

It is contemplated that one skilled in the art could use a phage display system in which fragments of the protein are expressed as fusion proteins on the phage surface (Pharmacia, Milwaukee WI). The recombinant DNA molecules are cloned into the phage DNA at a site which results in the transcription of a fusion protein a portion of which is encoded by the recombinant DNA molecule. The phage containing the recombinant nucleic acid molecule undergoes replication and transcription in the cell. The leader sequence of the fusion protein directs the transport of the fusion protein to the tip of the phage particle. Thus the fusion protein which is partially encoded by the recombinant DNA molecule is displayed on the phage particle for detection and selection by the methods described above.

It is further contemplated that a number of cycles of nucleic acid shuffling may be conducted with polynucleotides from a sub-population of the first population, which sub-population contains DNA encoding the desired recombinant protein. In this manner, proteins with even higher binding affinities or enzymatic activity could be achieved.

It is also contemplated that a number of cycles of nucleic acid shuffling may be conducted with a mixture of wild-type polynucleotides and a sub-population of nucleic acid from the first or subsequent rounds of nucleic acid shuffling in order to remove any silent mutations from the sub-population.

Any source of nucleic acid, in purified form can be utilized as the starting nucleic acid. Thus the process may employ DNA or RNA including messenger RNA, which DNA or RNA may be single or double stranded. In addition, a DNA-RNA hybrid which

contains one strand of each may be utilized. The nucleic acid sequence may be of various lengths depending on the size of the nucleic acid sequence to be mutated. Preferably the specific nucleic acid sequence is from 50 to 50000 base pairs. It is contemplated that entire vectors containing the nucleic acid encoding the protein of interest may be used in the methods of this invention.

The nucleic acid may be obtained from any source, for example, from plasmids such a pBR322, from cloned DNA or RNA or from natural DNA or RNA from any source including bacteria, yeast, viruses and higher organisms such as plants or animals. DNA or RNA may be extracted from blood or tissue material. The template polynucleotide may be obtained by amplification using the polynucleotide chain reaction (PCR, see **USPN 4,683,202** and **USPN 4,683,195**). Alternatively, the polynucleotide may be present in a vector present in a cell and sufficient nucleic acid may be obtained by culturing the cell and extracting the nucleic acid from the cell by methods known in the art.

Any specific nucleic acid sequence can be used to produce the population of hybrids by the present process. It is only necessary that a small population of hybrid sequences of the specific nucleic acid sequence exist or be created prior to the present process.

The initial small population of the specific nucleic acid sequences having mutations may be created by a number of different methods. Mutations may be created by error-prone PCR. Error-prone PCR uses low-fidelity polymerization conditions to introduce a low level of point mutations randomly over a long sequence. Alternatively, mutations can be introduced into the template polynucleotide by oligonucleotide-directed mutagenesis. In oligonucleotide-directed mutagenesis, a short sequence of the polynucleotide is removed from the polynucleotide using restriction enzyme digestion and is replaced with a synthetic polynucleotide in which various bases have been altered

from the original sequence. The polynucleotide sequence can also be altered by chemical mutagenesis. Chemical mutagens include, for example, sodium bisulfite, nitrous acid, hydroxylamine, hydrazine or formic acid. other agents which are analogues of nucleotide precursors include nitrosoguanidine, 5-bromouracil, 2-aminopurine, or acridine. Generally, these agents are added to the PCR reaction in place of the nucleotide precursor thereby mutating the sequence. Intercalating agents such as proflavine, acriflavine, quinacrine and the like can also be used. Random mutagenesis of the polynucleotide sequence can also be achieved by irradiation with X-rays or ultraviolet light. Generally, plasmid polynucleotides so mutagenized are introduced into *E. coli* and propagated as a pool or library of hybrid plasmids.

Alternatively the small mixed population of specific nucleic acids may be found in nature in that they may consist of different alleles of the same gene or the same gene from different related species (i.e., cognate genes). Alternatively, they may be related DNA sequences found within one species, for example, the immunoglobulin genes.

Once the mixed population of the specific nucleic acid sequences is generated, the polynucleotides can be used directly or inserted into an appropriate cloning vector, using techniques well-known in,the art.

The choice of vector depends on the size of the polynucleotide sequence and the host cell to be employed in the methods of this invention. The templates of this invention may be plasmids, phages, cosmids, phagemids, viruses (e.g., retroviruses, parainfluenzavirus, herpesviruses, reoviruses, paramyxoviruses, and the like), or selected portions thereof (e.g., coat protein, spike glycoprotein, capsid protein). For example, cosmids and phagemids are preferred where the specific nucleic acid sequence to be mutated is larger because these vectors are able to stably propagate large polynucleotides.

If the mixed population of the specific nucleic acid sequence is cloned into a vector it can be clonally amplified by inserting each vector into a host cell and allowing the host cell to amplify the vector. This is referred to as clonal amplification because while the absolute number of nucleic acid sequences increases, the number of hybrids does not increase. Utility can be readily determined by screening expressed polypeptides.

The DNA shuffling method of this invention can be performed blindly on a pool of unknown sequences. By adding to the reassembly mixture oligonucleotides (with ends that are homologous to the sequences being reassembled) any sequence mixture can be incorporated at any specific position into another sequence mixture. Thus, it is contemplated that mixtures of synthetic oligonucleotides, PCR polynucleotides or even whole genes can be mixed into another sequence library at defined positions. The insertion of one sequence (mixture) is independent from the insertion of a sequence in another part of the template. Thus, the degree of recombination, the homology required, and the diversity of the library can be independently and simultaneously varied along the length of the reassembled DNA.

This approach of mixing two genes may be useful for the humanization of antibodies from murine hybridomas. The approach of mixing two genes or inserting alternative sequences into genes may be useful for any therapeutically used protein, for example, interleukin I, antibodies, tPA and growth hormone. The approach may also be useful in any nucleic acid for example, promoters or introns or 31 untranslated region or 51 untranslated regions of genes to increase expression or alter specificity of expression of proteins. The approach may also be used to mutate ribozymes or aptamers.

Shuffling requires the presence of homologous regions separating regions of diversity. Scaffold-like protein structures may be particularly suitable for shuffling. The conserved scaffold determines the overall folding by self-association, while displaying relatively unrestricted loops that mediate the specific binding. Examples of such

scaffolds are the immunoglobulin beta-barrel, and the four-helix bundle which are well-known in the art. This shuffling can be used to create scaffold-like proteins with various combinations of mutated sequences for binding.

5

***In vitro* Shuffling**

The equivalents of some standard genetic matings may also be performed by shuffling *in vitro*. For example, a "molecular backcross" can be performed by repeatedly 10 mixing the hybrid's nucleic acid with the wild-type nucleic acid while selecting for the mutations of interest. As in traditional breeding, this approach can be used to combine phenotypes from different sources into a background of choice. It is useful, for example, for the removal of neutral mutations that affect unselected characteristics (i.e. immunogenicity). Thus it can be useful to determine which mutations in a protein are 15 involved in the enhanced biological activity and which are not, an advantage which cannot be achieved by error-prone mutagenesis or cassette mutagenesis methods.

Large, functional genes can be assembled correctly from a mixture of small random polynucleotides. This reaction may be of use for the reassembly of genes from 20 the highly fragmented DNA of fossils. In addition random nucleic acid fragments from fossils may be combined with polynucleotides from similar genes from related species.

It is also contemplated that the method of this invention can be used for the *in vitro* amplification of a whole genome from a single cell as is needed for a variety of 25 research and diagnostic applications. DNA amplification by PCR is in practice limited to a length of about 40 kb. Amplification of a whole genome such as that of *E.* coli (5, 000 kb) by PCR would require about 250 primers yielding 125 forty kb polynucleotides. This approach is not practical due to the unavailability of sufficient sequence data. On the other hand, random production of polynucleotides of the genome with sexual PCR cycles, 30 followed by gel purification of small polynucleotides will provide a multitude of possible

primers. Use of this mix of random small polynucleotides as primers in a PCR reaction alone or with the whole genome as the template should result in an inverse chain reaction with the theoretical endpoint of a single concatamer containing many copies of the genome.

100 fold amplification in the copy number and an average polynucleotide size of greater than 50 kb may be obtained when only random polynucleotides are used. It is thought that the larger concatamer is generated by overlap of many smaller polynucleotides. The quality of specific PCR products obtained using synthetic primers will be indistinguishable from the product obtained from unamplified DNA. It is expected that this approach will be useful for the mapping of genomes.

The polynucleotide to be shuffled can be produced as random or non-random polynucleotides, at the discretion of the practitioner. Moreover, this invention provides a method of shuffling that is applicable to a wide range of polynucleotide sizes and types, including the step of generating polynucleotide monomers to be used as building blocks in the reassembly of a larger polynucleotide. For example, the building blocks can be fragments of genes or they can be comprised of entire genes or gene pathways, or any combination thereof.

In vivo Shuffling

In an embodiment of *in vivo* shuffling, the mixed population of the specific nucleic acid sequence is introduced into bacterial or eukaryotic cells under conditions such that at least two different nucleic acid sequences are present in each host cell. The polynucleotides can be introduced into the host cells by a variety of different methods. The host cells can be transformed with the smaller polynucleotides using methods known in the art, for example treatment with calcium chloride. If the polynucleotides are inserted into a phage genome, the host cell can be transfected with the recombinant phage

genome having the specific nucleic acid sequences. Alternatively, the nucleic acid sequences can be introduced into the host cell using electroporation, transfection, lipofection, biolistics, conjugation, and the like.

In general, in this embodiment, the specific nucleic acids sequences will be present in vectors which are capable of stably replicating the sequence in the host cell. In addition, it is contemplated that the vectors will encode a marker gene such that host cells having the vector can be selected. This ensures that the mutated specific nucleic acid sequence can be recovered after introduction into the host cell. However, it is contemplated that the entire mixed population of the specific nucleic acid sequences need not be present on a vector sequence. Rather only a sufficient number of sequences need be cloned into vectors to ensure that after introduction of the polynucleotides into the host cells each host cell contains one vector having at least one specific nucleic acid sequence present therein. It is also contemplated that rather than having a subset of the population of the specific nucleic acids sequences cloned into vectors, this subset may be already stably integrated into the host cell.

It has been found that when two polynucleotides which have regions of identity are inserted into the host cells homologous recombination occurs between the two polynucleotides. Such recombination between the two mutated specific nucleic acid sequences will result in the production of double or triple hybrids in some situations.

It has also been found that the frequency of recombination is increased if some of the mutated specific nucleic acid sequences are present on linear nucleic acid molecules. Therefore, in a preferred embodiment, some of the specific nucleic acid sequences are present on linear polynucleotides.

After transformation, the host cell transformants are placed under selection to identify those host cell transformants which contain mutated specific nucleic acid

sequences having the qualities desired. For example, if increased resistance to a particular drug is desired then the transformed host cells may be subjected to increased concentrations of the particular drug and those transformants producing mutated proteins able to confer increased drug resistance will be selected. If the enhanced ability of a particular protein to bind to a receptor is desired, then expression of the protein can be induced from the transformants and the resulting protein assayed in a ligand binding assay by methods known in the art to identify that subset of the mutated population which shows enhanced binding to the ligand. Alternatively, the protein can be expressed in another system to ensure proper processing.

Once a subset of the first recombined specific nucleic acid sequences (daughter sequences) having the desired characteristics are identified, they are then subject to a second round of recombination.

In the second cycle of recombination, the recombined specific nucleic acid sequences may be mixed with the original mutated specific nucleic acid sequences (parent sequences) and the cycle repeated as described above. In this way a set of second recombined specific nucleic acids sequences can be identified which have enhanced characteristics or encode for proteins having enhanced properties. This cycle can be repeated a number of times as desired.

It is also contemplated that in the second or subsequent recombination cycle, a backcross can be performed. A molecular backcross can be performed by mixing the desired specific nucleic acid sequences with a large number of the wild-type sequence, such that at least one wild-type nucleic acid sequence and a mutated nucleic acid sequence are present in the same host cell after transformation. Recombination with the wild-type specific nucleic acid sequence will eliminate those neutral mutations that may affect unselected characteristics such as immunogenicity but not the selected characteristics.

In another embodiment of this invention, it is contemplated that during the first round a subset of the specific nucleic acid sequences can be generated as smaller polynucleotides by slowing or halting their PCR amplification prior to introduction into the host cell. The size of the polynucleotides must be large enough to contain some regions of identity with the other sequences so as to homologously recombine with the other sequences. The size of the polynucleotides will range from 0.03 kb to 100 kb more preferably from 0. 2 kb to 10 kb. It is also contemplated that in subsequent rounds, all of the specific nucleic acid sequences other than the sequences selected from the previous round may be utilized to generate PCR polynucleotides prior to introduction into the host cells.

The shorter polynucleotide sequences can be single-stranded or double-stranded. If the sequences were originally single-stranded and have become double-stranded they can be denatured with heat, chemicals or enzymes prior to insertion into the host cell. The reaction conditions suitable for separating the strands of nucleic acid are well known in the art.

The steps of this process can be repeated indefinitely, being limited only by the number of possible hybrids which can be achieved. After a certain number of cycles, all possible hybrids will have been achieved and further cycles are redundant.

In an embodiment the same mutated template nucleic acid is repeatedly recombined and the resulting recombinants selected for the desired characteristic.

Therefore, the initial pool or population of mutated template nucleic acid is cloned into a vector capable of replicating in a bacteria such as *E. coli.* The particular vector is not essential, so long as it is capable of autonomous replication in E. coli. In a preferred embodiment, the vector is designed to allow the expression and production of

any protein encoded by the mutated specific nucleic acid linked to the vector. It is also preferred that the vector contain a gene encoding for a selectable marker.

The population of vectors containing the pool of mutated nucleic acid sequences is introduced into the *E. coli* host cells. The vector nucleic acid sequences may be introduced by transformation, transfection or infection in the case of phage. The concentration of vectors used to transform the bacteria is such that a number of vectors is introduced into each cell. Once present in the cell, the efficiency of homologous recombination is such that homologous recombination occurs between the various vectors. This results in the generation of hybrids (daughters) having a combination of mutations which differ from the original parent mutated sequences.

The host cells are then clonally replicated and selected for the marker gene present on the vector. Only those cells having a plasmid will grow under the selection.

The host cells which contain a vector are then tested for the presence of favorable mutations. Such testing may consist of placing the cells under selective pressure, for example, if the gene to be selected is an improved drug resistance gene. If the vector allows expression of the protein encoded by the mutated nucleic acid sequence, then such selection may include allowing expression of the protein so encoded, isolation of the protein and testing of the protein to determine whether, for example, it binds with increased efficiency to the ligand of interest.

Once a particular daughter mutated nucleic acid sequence has been identified which confers the desired characteristics, the nucleic acid is isolated either already linked to the vector or separated from the vector. This nucleic acid is then mixed with the first or parent population of nucleic acids and the cycle is repeated.

It has been shown that by this method nucleic acid sequences having enhanced desired properties can be selected.

In an alternate embodiment, the first generation of hybrids are retained in the cells and the parental mutated sequences are added again to the cells. Accordingly, the first cycle of Embodiment I is conducted as described above. However, after the daughter nucleic acid sequences are identified, the host cells containing these sequences are retained.

The parent mutated specific nucleic acid population, either as polynucleotides or cloned into the same vector is introduced into the host cells already containing the daughter nucleic acids. Recombination is allowed to occur in the cells and the next generation of recombinants, or granddaughters are selected by the methods described above.

This cycle can be repeated a number of times until the nucleic acid or peptide having the desired characteristics is obtained. It is contemplated that in subsequent cycles, the population of mutated sequences which are added to the preferred hybrids may come from the parental hybrids or any subsequent generation.

In an alternative embodiment, the invention provides a method of conducting a "molecular" backcross of the obtained recombinant specific nucleic acid in order to eliminate any neutral mutations. Neutral mutations are those mutations which do not confer onto the nucleic acid or peptide the desired properties. Such mutations may however confer on the nucleic acid or peptide undesirable characteristics. Accordingly, it is desirable to eliminate such neutral mutations. The method of this invention provide a means of doing so.

In this embodiment, after the hybrid nucleic acid, having the desired characteristics, is obtained by the methods of the embodiments, the nucleic acid, the

vector having the nucleic acid or the host cell containing the vector and nucleic acid is isolated.

The nucleic acid or vector is then introduced into the host cell with a large excess of the wild-type nucleic acid. The nucleic acid of the hybrid and the nucleic acid of the wild-type sequence are allowed to recombine. The resulting recombinants are placed under the same selection as the hybrid nucleic acid. Only those recombinants which retained the desired characteristics will be selected. Any silent mutations which do not provide the desired characteristics will be lost through recombination with the wild-type DNA. This cycle can be repeated a number of times until all of the silent mutations are eliminated.

Thus the methods of this invention can be used in a molecular backcross to eliminate unnecessary or silent mutations.

2.11.2.3. EXONUCLEASE-MEDIATED REASSEMBLY

In a particular embodiment, this invention provides for a method for shuffling, assembling, reassembling, recombining, &/or concatenating at least two polynucleotides to form a progeny polynucleotide (e.g. a chimeric progeny polynucleotide that can be expressed to produce a polypeptide or a gene pathway). In a particular embodiment, a double stranded polynucleotide end (e.g. two single stranded sequences hybridized to each other as hybridization partners) is treated with an exonuclease to liberate nucleotides from one of the two strands, leaving the remaining strand free of its original partner so that, if desired, the remaining strand may be used to achieve hybridization to another partner.

In a particular aspect, a double stranded polynucleotide end (that may be part of - or connected to - a polynucleotide or a nonpolynucleotide sequence) is subjected to a

source of exonuclease activity. Serviceable sources of exonuclease activity may be an enzyme with 3' exonuclease activity, an enzyme with 5' exonuclease activity, an enzyme with both 3' exonuclease activity and 5' exonuclease activity, and any combination thereof. An exonuclease can be used to liberate nucleotides from one or both ends of a linear double stranded polynucleotide, and from one to all ends of a branched polynucleotide having more than two ends. The mechanism of action of this liberation is believed to be comprised of an enzymatically-catalyzed hydrolysis of terminal nucleotides, and can be allowed to proceed in a time-dependent fashion, allowing experimental control of the progression of the enzymatic process.

By contrast, a non-enzymatic step may be used to shuffle, assemble, reassemble, recombine, and/or concatenate polynucleotide building blocks that is comprised of subjecting a working sample to denaturing (or "melting") conditions (for example, by changing temperature, pH, and /or salinity conditions) so as to melt a working set of double stranded polynucleotides into single polynucleotide strands. For shuffling, it is desirable that the single polynucleotide strands participate to some extent in annealment with different hybridization partners (i.e. and not merely revert to exclusive reannealment between what were former partners before the denaturation step). The presence of the former hybridization partners in the reaction vessel, however, does not preclude, and may sometimes even favor, reannealment of a single stranded polynucleotide with its former partner, to recreate an original double stranded polynucleotide.

In contrast to this non-enzymatic shuffling step comprised of subjecting double stranded polynucleotide building blocks to denaturation, followed by annealment, the instant invention further provides an exonuclease-based approach requiring no denaturation – rather, the avoidance of denaturing conditions and the maintenance of double stranded polynucleotide substrates in annealed (i.e. non-denatured) state are necessary conditions for the action of exonucleases (e.g., exonuclease III and red alpha gene product). Additionally in contrast, the generation of single stranded polynucleotide

sequences capable of hybridizing to other single stranded polynucleotide sequences is the result of covalent cleavage – and hence sequence destruction - in one of the hybridization partners. For example, an exonuclease III enzyme may be used to enzymatically liberate 3' terminal nucleotides in one hybridization strand (to achieve covalent hydrolysis in that polynucleotide strand); and this favors hybridization of the remaining single strand to a new partner (since its former partner was subjected to covalent cleavage).

By way of further illustration, a specific exonuclease, namely exonuclease III is provided herein as an example of a 3' exonuclease; however, other exonucleases may also be used, including enzymes with 5' exonuclease activity and enzymes with 3' exonuclease activity, and including enzymes not yet discovered and enzymes not yet developed. It is particularly appreciated that enzymes can be discovered, optimized (e.g. engineered by directed evolution), or both discovered and optimized specifically for the instantly disclosed approach that have more optimal rates &/or more highly specific activities &/or greater lack of unwanted activities. In fact it is expected that the instant invention may encourage the discovery &/or development of such designer enzymes. In sum, this invention may be practiced with a variety of currently available exonuclease enzymes, as well enzymes not yet discovered and enzymes not yet developed.

The exonuclease action of exonuclease III requires a working double stranded polynucleotide end that is either blunt or has a 5' overhang, and the exonuclease action is comprised of enzymatically liberating 3' terminal nucleotides, leaving a single stranded 5' end that becomes longer and longer as the exonuclease action proceeds (see Figure 1). Any 5' overhangs produced by this approach may be used to hybridize to another single stranded polynucleotide sequence (which may also be a single stranded polynucleotide or a terminal overhang of a partially double stranded polynucleotide) that shares enough homology to allow hybridization. The ability of these exonuclease III-generated single stranded sequences (e.g. in 5' overhangs) to hybridize to other single stranded sequences

allows two or more polynucleotides to be shuffled, assembled, reassembled, &/or concatenated.

Furthermore, it is appreciated that one can protect the end of a double stranded polynucleotide or render it susceptible to a desired enzymatic action of a serviceable exonuclease as necessary. For example, a double stranded polynucleotide end having a 3' overhang is not susceptible to the exonuclease action of exonuclease III. However, it may be rendered susceptible to the exonuclease action of exonuclease III by a variety of means; for example, it may be blunted by treatment with a polymerase, cleaved to provide a blunt end or a 5' overhang, joined (ligated or hybridized) to another double stranded polynucleotide to provide a blunt end or a 5' overhang, hybridized to a single stranded polynucleotide to provide a blunt end or a 5' overhang, or modified by any of a variety of means).

According to one aspect, an exonuclease may be allowed to act on one or on both ends of a linear double stranded polynucleotide and proceed to completion, to near completion, or to partial completion. When the exonuclease action is allowed to go to completion, the result will be that the length of each 5' overhang will be extend far towards the middle region of the polynucleotide in the direction of what might be considered a "rendezvous point" (which may be somewhere near the polynucleotide midpoint). Ultimately, this results in the production of single stranded polynucleotides (that can become dissociated) that are each about half the length of the original double stranded polynucleotide (see Figure 1). Alternatively, an exonuclease-mediated reaction can be terminated before proceeding to completion.

Thus this exonuclease-mediated approach is serviceable for shuffling, assembling &/or reassembling, recombining, and concatenating polynucleotide building blocks, which polynucleotide building blocks can be up to ten bases long or tens of bases long or

hundreds of bases long or thousands of bases long or tens of thousands of bases long or hundreds of thousands of bases long or millions of bases long or even longer.

This exonuclease-mediated approach is based on the action of double stranded DNA specific exodeoxyribonuclease activity of *E. coli* exonuclease III. Substrates for exonuclease III may be generated by subjecting a double stranded polynucleotide to fragmentation. Fragmentation may be achieved by mechanical means (e.g., shearing, sonication, etc.), by enzymatic means (e.g. using restriction enzymes), and by any combination thereof. Fragments of a larger polynucleotide may also be generated by polymerase-mediated synthesis.

Exonuclease III is a 28K monomeric enzyme, product of the *xth*A gene of *E. coli* with four known activities: exodeoxyribonuclease (alternatively referred to as exonuclease herein), RNaseH, DNA-3'-phosphatase, and AP endonuclease. The exodeoxyribonuclease activity is specific for double stranded DNA. The mechanism of action is thought to involve enzymatic hydrolysis of DNA from a 3' end progressively towards a 5' direction, with formation of nucleoside 5'-phosphates and a residual single strand. The enzyme does not display efficient hydrolysis of single stranded DNA, single-stranded RNA, or double-stranded RNA; however it degrades RNA in an DNA-RNA hybrid releasing nucleoside 5'-phosphates. The enzyme also releases inorganic phosphate specifically from 3'phosphomonoester groups on DNA, but not from RNA or short oligonucleotides. Removal of these groups converts the terminus into a primer for DNA polymerase action.

Additional examples of enzymes with exonuclease activity include red-alpha and venom phosphodiesterases. Red alpha (*redα*) gene product (also referred to as lambda exonuclease) is of bacteriophage λ origin. The *redα* gene is transcribed from the leftward promoter and its product is involved (24 kD) in recombination. Red alpha gene product acts processively from 5'-phosphorylated termini to liberate mononucleotides from

duplex DNA (**Takahashi & Kobayashi**, 1990). Venom phosphodiesterases (Laskowski, 1980) is capable of rapidly opening supercoiled DNA.

2.11.2.3. NON-STOCHASTIC LIGATION REASSEMBLY

In one aspect, the present invention provides a non-stochastic method termed synthetic ligation reassembly (SLR), that is somewhat related to stochastic shuffling, save that the nucleic acid building blocks are not shuffled or concatenated or chimerized randomly, but rather are assembled non-stochastically.

A particularly glaring difference is that the instant SLR method does not depend on the presence of a high level of homology between polynucleotides to be shuffled. In contrast, prior methods, particularly prior stochastic shuffling methods require that presence of a high level of homology, particularly at coupling sites, between polynucleotides to be shuffled. Accordingly these prior methods favor the regeneration of the original progenitor molecules, and are suboptimal for generating large numbers of novel progeny chimeras, particularly full-length progenies. The instant invention, on the other hand, can be used to non-stochastically generate libraries (or sets) of progeny molecules comprised of over 10^{100} different chimeras. Conceivably, SLR can even be used to generate libraries comprised of over 10^{1000} different progeny chimeras with (no upper limit in sight).

Thus, in one aspect, the present invention provides a method, which method is non-stochastic, of producing a set of finalized chimeric nucleic acid molecules having an overall assembly order that is chosen by design, which method is comprised of the steps of generating by design a plurality of specific nucleic acid building blocks having serviceable mutually compatible ligatable ends, and assembling these nucleic acid building blocks, such that a designed overall assembly order is achieved.

The mutually compatible ligatable ends of the nucleic acid building blocks to be assembled are considered to be "serviceable" for this type of ordered assembly if they enable the building blocks to be coupled in predetermined orders. Thus, in one aspect, the overall assembly order in which the nucleic acid building blocks can be coupled is specified by the design of the ligatable ends and, if more than one assembly step is to be used, then the overall assembly order in which the nucleic acid building blocks can be coupled is also specified by the sequential order of the assembly step(s). Figure 4, Panel C illustrates an exemplary assembly process comprised of 2 sequential steps to achieve a designed (non-stochastic) overall assembly order for five nucleic acid building blocks. In a preferred embodiment of this invention, the annealed building pieces are treated with an enzyme, such as a ligase (e.g. T4 DNA ligase), achieve covalent bonding of the building pieces.

In a preferred embodiment, the design of nucleic acid building blocks is obtained upon analysis of the sequences of a set of progenitor nucleic acid templates that serve as a basis for producing a progeny set of finalized chimeric nucleic acid molecules. These progenitor nucleic acid templates thus serve as a source of sequence information that aids in the design of the nucleic acid building blocks that are to be mutagenized, i.e. chimerized or shuffled.

In one exemplification, this invention provides for the chimerization of a family of related genes and their encoded family of related products. In a particular exemplification, the encoded products are enzymes. As a representative list of families of enzymes which may be mutagenized in accordance with the aspects of the present invention, there may be mentioned, the following enzymes and their functions:

1 **Lipase/Esterase**

a. Enantioselective hydrolysis of esters (lipids)/ thioesters

1) Resolution of racemic mixtures

2) Synthesis of optically active acids or alcohols from *meso*-diesters

b. Selective syntheses

1) Regiospecific hydrolysis of carbohydrate esters

2) Selective hydrolysis of cyclic secondary alcohols

c. Synthesis of optically active esters, lactones, acids, alcohols

1) Transesterification of activated/nonactivated esters

2) Interesterification

3) Optically active lactones from hydroxyesters

4) Regio- and enantioselective ring opening of anhydrides

d. Detergents

e. Fat/Oil conversion

f. Cheese ripening

2 **Protease**

a. Ester/amide synthesis

b. Peptide synthesis

c. Resolution of racemic mixtures of amino acid esters

d. Synthesis of non-natural amino acids

e. Detergents/protein hydrolysis

3 **Glycosidase/Glycosyl transferase**

a. Sugar/polymer synthesis

b. Cleavage of glycosidic linkages to form mono, di-and oligosaccharides

c. Synthesis of complex oligosaccharides

d. Glycoside synthesis using UDP-galactosyl transferase

e. Transglycosylation of disaccharides, glycosyl fluorides, aryl galactosides

f. Glycosyl transfer in oligosaccharide synthesis

g. Diastereoselective cleavage of β-glucosylsulfoxides

h. Asymmetric glycosylations

i. Food processing

j. Paper processing

4 **Phosphatase/Kinase**

a. Synthesis/hydrolysis of phosphate esters

1) Regio-, enantioselective phosphorylation

2) Introduction of phosphate esters

3) Synthesize phospholipid precursors

4) Controlled polynucleotide synthesis

b. Activate biological molecule

c. Selective phosphate bond formation without protecting groups

5 **Mono/Dioxygenase**

a. Direct oxyfunctionalization of unactivated organic substrates

b. Hydroxylation of alkane, aromatics, steroids

c. Epoxidation of alkenes

d. Enantioselective sulphoxidation

e. Regio- and stereoselective Bayer-Villiger oxidations

6 **Haloperoxidase**

a. Oxidative addition of halide ion to nucleophilic sites

b. Addition of hypohalous acids to olefinic bonds

c. Ring cleavage of cyclopropanes

d. Activated aromatic substrates converted to *ortho* and *para* derivatives

e. 1.3 diketones converted to 2-halo-derivatives

f. Heteroatom oxidation of sulfur and nitrogen containing substrates

g. Oxidation of enol acetates, alkynes and activated aromatic rings

7 **Lignin peroxidase/Diarylpropane peroxidase**

a. Oxidative cleavage of C-C bonds

b. Oxidation of benzylic alcohols to aldehydes

c. Hydroxylation of benzylic carbons

d. Phenol dimerization

e. Hydroxylation of double bonds to form diols

f. Cleavage of lignin aldehydes

8 **Epoxide hydrolase**

a. Synthesis of enantiomerically pure bioactive compounds

b. Regio- and enantioselective hydrolysis of epoxide

c. Aromatic and olefinic epoxidation by monooxygenases to form epoxides

d. Resolution of racemic epoxides

e. Hydrolysis of steroid epoxides

9 **Nitrile hydratase/nitrilase**

a. Hydrolysis of aliphatic nitriles to carboxamides

b. Hydrolysis of aromatic, heterocyclic, unsaturated aliphatic nitriles to corresponding acids

c. Hydrolysis of acrylonitrile

d. Production of aromatic and carboxamides, carboxylic acids (nicotinamide, picolinamide, isonicotinamide)

e. Regioselective hydrolysis of acrylic dinitrile

f. α-amino acids from α-hydroxynitriles

10 **Transaminase**

a. Transfer of amino groups into oxo-acids

11 **Amidase/Acylase**

a. Hydrolysis of amides, amidines, and other C-N bonds

b. Non-natural amino acid resolution and synthesis

These exemplifications, while illustrating certain specific aspects of the invention, do not portray the limitations or circumscribe the scope of the disclosed invention.

Thus according to one aspect of this invention, the sequences of a plurality of progenitor nucleic acid templates are aligned in order to select one or more demarcation points, which demarcation points can be located at an area of homology, and are comprised of one or more nucleotides, and which demarcation points are shared by at least two of the progenitor templates. The demarcation points can be used to delineate the boundaries of nucleic acid building blocks to be generated. Thus, the demarcation points identified and selected in the progenitor molecules serve as potential chimerization points in the assembly of the progeny molecules.

Preferably a serviceable demarcation point is an area of homology (comprised of at least one homologous nucleotide base) shared by at least two progenitor templates. More preferably a serviceable demarcation point is an area of homology that is shared by at least half of the progenitor templates. More preferably still a serviceable demarcation point is an area of homology that is shared by at least two thirds of the progenitor templates. Even more preferably a serviceable demarcation points is an area of homology that is shared by at least three fourths of the progenitor templates. Even more preferably still a serviceable demarcation points is an area of homology that is shared by

at almost all of the progenitor templates. Even more preferably still a serviceable demarcation point is an area of homology that is shared by all of the progenitor templates.

The process of designing nucleic acid building blocks and of designing the mutually compatible ligatable ends of the nucleic acid building blocks to be assembled is illustrated in Figures 6 and 7. As shown, the alignment of a set of progenitor templates reveals several naturally occurring demarcation points, and the identification of demarcation points shared by these templates helps to non-stochastically determine the building blocks to be generated and used for the generation of the progeny chimeric molecules.

In a preferred embodiment, this invention provides that the ligation reassembly process is performed exhaustively in order to generate an exhaustive library. In other words, all possible ordered combinations of the nucleic acid building blocks are represented in the set of finalized chimeric nucleic acid molecules. At the same time, in a particularly preferred embodiment, the assembly order (i.e. the order of assembly of each building block in the 5' to 3 sequence of each finalized chimeric nucleic acid) in each combination is by design (or non-stochastic). Because of the non-stochastic nature of this invention, the possibility of unwanted side products is greatly reduced.

In another preferred embodiment, this invention provides that, the ligation reassembly process is performed systematically, for example in order to generate a systematically compartmentalized library, with compartments that can be screened systematically, e.g. one by one. In other words this invention provides that, through the selective and judicious use of specific nucleic acid building blocks, coupled with the selective and judicious use of sequentially stepped assembly reactions, an experimental design can be achieved where specific sets of progeny products are made in each of several reaction vessels. This allows a systematic examination and screening procedure

to be performed. Thus, it allows a potentially very large number of progeny molecules to be examined systematically in smaller groups.

Because of its ability to perform chimerizations in a manner that is highly flexible yet exhaustive and systematic as well, particularly when there is a low level of homology among the progenitor molecules, the instant invention provides for the generation of a library (or set) comprised of a large number of progeny molecules. Because of the non-stochastic nature of the instant ligation reassembly invention, the progeny molecules generated preferably comprise a library of finalized chimeric nucleic acid molecules having an overall assembly order that is chosen by design. In a particularly preferred embodiment of this invention, such a generated library is comprised of preferably greater than 10^3 different progeny molecular species, more preferably greater than 10^5 different progeny molecular species, more preferably still greater than 10^{10} different progeny molecular species, more preferably still greater than 10^{15} different progeny molecular species, more preferably still greater than 10^{20} different progeny molecular species, more preferably still greater than 10^{30} different progeny molecular species, more preferably still greater than 10^{40} different progeny molecular species, more preferably still greater than 10^{50} different progeny molecular species, more preferably still greater than 10^{60} different progeny molecular species, more preferably still greater than 10^{70} different progeny molecular species, more preferably still greater than 10^{80} different progeny molecular species, more preferably still greater than 10^{100} different progeny molecular species, more preferably still greater than 10^{110} different progeny molecular species, more preferably still greater than 10^{120} different progeny molecular species, more preferably still greater than 10^{130} different progeny molecular species, more preferably still greater than 10^{140} different progeny molecular species, more preferably still greater than 10^{150} different progeny molecular species, more preferably still greater than 10^{175} different progeny molecular species, more preferably still greater than 10^{200} different progeny molecular species, more preferably still greater than 10^{300} different progeny molecular species, more preferably still greater than 10^{400} different progeny molecular species, more

preferably still greater than 10^{500} different progeny molecular species, and even more preferably still greater than 10^{1000} different progeny molecular species.

In one aspect, a set of finalized chimeric nucleic acid molecules, produced as described is comprised of a polynucleotide encoding a polypeptide. According to one preferred embodiment, this polynucleotide is a gene, which may be a man-made gene. According to another preferred embodiment, this polynucleotide is a gene pathway, which may be a man-made gene pathway. This invention provides that one or more man-made genes generated by this invention may be incorporated into a man-made gene pathway, such as pathway operable in a eukaryotic organism (including a plant).

It is appreciated that the power of this invention is exceptional, as there is much freedom of choice and control regarding the selection of demarcation points, the size and number of the nucleic acid building blocks, and the size and design of the couplings. It is appreciated, furthermore, that the requirement for intermolecular homology is highly relaxed for the operability of this invention. In fact, demarcation points can even be chosen in areas of little or no intermolecular homology. For example, because of codon wobble, i.e. the degeneracy of codons, nucleotide substitutions can be introduced into nucleic acid building blocks without altering the amino acid originally encoded in the corresponding progenitor template. Alternatively, a codon can be altered such that the coding for an originally amino acid is altered. This inventiop provides that such substitutions can be introduced into the nucleic acid building block in order to increase the incidence of intermolecularly homologous demarcation points and thus to allow an increased number of couplings to be achieved among the building blocks, which in turn allows a greater number of progeny chimeric molecules to be generated.

In another exemplifaction, the synthetic nature of the step in which the building blocks are generated allows the design and introduction of nucleotides (e.g. one or more nucleotides, which may be, for example, codons or introns or regulatory sequences) that

can later be optionally removed in an in vitro process (e.g. by mutageneis) or in an in vivo process (e.g. by utilizing the gene splicing ability of a host organism). It is appreciated that in many instances the introduction of these nucleotides may also be desirable for many other reasons in addition to the potential benefit of creating a serviceable demarcation point.

Thus, according to another embodiment, this invention provides that a nucleic acid building block can be used to introduce an intron. Thus, this invention provides that functional introns may be introduced into a man-made gene of this invention. This invention also provides that functional introns may be introduced into a man-made gene pathway of this invention. Accordingly, this invention provides for the generation of a chimeric polynucleotide that is a man-made gene containing one (or more) artificially introduced intron(s).

Accordingly, this invention also provides for the generation of a chimeric polynucleotide that is a man-made gene pathway containing one (or more) artificially introduced intron(s). Preferably, the artificially introduced intron(s) are functional in one or more host cells for gene splicing much in the way that naturally-occurring introns serve functionally in gene splicing. This invention provides a process of producing man-made intron-containing polynucleotides to be introduced into host organisms for recombination and/or splicing.

The ability to achieve chimerizations, using couplings as described herein, in areas of little or no homology among the progenitor molecules, is particularly useful, and in fact critical, for the assembly of novel gene pathways. This invention thus provides for the generation of novel man-made gene pathways using synthetic ligation reassembly. In a particular aspect, this is achieved by the introduction of regulatory sequences, such as promoters, that are operable in an intended host, to confer operability to a novel gene pathway when it is introduced into the intended host. In a particular exemplification, this

invention provides for the generation of novel man-made gene pathways that is operable in a plurality of intended hosts (e.g. in a microbial organism as well as in a plant cell). This can be achieve, for example, by the introduction of a plurality of regulatory sequences, comprised of a regulatory sequence that is operable in a first intended host and a regulatory sequence that is operable in a second intended host. A similar process can be performed to achieve operability of a gene pathway in a third intended host species, etc. The number of intended host species can be each integer from 1 to 10 or alternatively over 10. Alternatively, for example, operability of a gene pathway in a plurality of intended hosts can be achieved by the introduction of a regulatory sequence having intrinsic operability in a plurality of intended hosts.

Thus, according to a particular embodiment, this invention provides that a nucleic acid building block can be used to introduce a regulatory sequence, particularly a regulatory sequence for gene expression. Preferred regulatory sequences include, but are not limited to, those that are man-made, and those found in archeal, bacterial, eukaryotic (including mitochondrial), viral, and prionic or prion-like organisms. Preferred regulatory sequences include but are not limited to, promoters, operators, and activator binding sites. Thus, this invention provides that functional regulatory sequences may be introduced into a man-made gene of this invention. This invention also provides that functional regulatory sequences may be introduced into a man-made gene pathway of this invention.

Accordingly, this invention provides for the generation of a chimeric polynucleotide that is a man-made gene containing one (or more) artificially introduced regulatory sequence(s). Accordingly, this invention also provides for the generation of a chimeric polynucleotide that is a man-made gene pathway containing one (or more) artificially introduced regulatory sequence(s). Preferably, an artificially introduced regulatory sequence(s) is operatively linked to one or more genes in the man-made polynucleotide, and are functional in one or more host cells.

Preferred bacterial promoters that are serviceable for this invention include lacI, lacZ, T3, T7, gpt, lambda P_R, P_L and trp. Serviceable eukaryotic promoters include CMV immediate early, HSV thymidine kinase, early and late SV40, LTRs from retrovirus, and mouse metallothionein-I. Particular plant regulatory sequences include promoters active in directing transcription in plants, either constitutively or stage and/or tissue specific, depending on the use of the plant or parts thereof. These promoters include, but are not limited to promoters showing constitutive expression, such as the 35S promoter of Cauliflower Mosaic Virus (CaMV) (**Guilley** et al., 1982), those for leaf-specific expression, such as the promoter of the ribulose bisphosphate carboxylase small subunit gene (**Coruzzi** et al., 1984), those for root-specific expression, such as the promoter from the glutamin synthase gene (**Tingey** et al., 1987), those for seed-specific expression, such as the cruciferin A promoter from *Brassica napus* (**Ryan** et al., 1989), those for tuber-specific expression, such as the class-I patatin promoter from potato (**Rocha-Sasa** et al., 1989; **Wenzler** et al., 1989) or those for fruit-specific expression, such as the polygalacturonase (PG) promoter from tomato (**Bird** et al., 1988).

Other regulatory sequences that are preferred for this invention include terminator sequences and polyadenylation signals and any such sequence functioning as such in plants, the choice of which is within the level of the skilled artisan. An example of such sequences is the 3′ flanking region of the nopaline synthase (nos) gene of *Agrobacterium tumefaciens* (**Bevan**, 1984). The regulatory sequences may also include enhancer sequences, such as found in the 35S promoter of CaMV, and mRNA stabilizing sequences such as the leader sequence of Alfalfa Mosaic Cirus (AlMV) RNA4 (**Brederode** et al., 1980) or any other sequences functioning in a like manner.

A man-made genes produced using this invention can also serve as a substrate for recombination with another nucleic acid. Likewise, a man-made gene pathway produced using this invention can also serve as a substrate for recombination with another nucleic

acid. In a preferred instance, the recombination is facilitated by, or occurs at, areas of homology between the man-made intron-containing gene and a nucleic acid with serves as a recombination partner. In a particularly preferred instance, the recombination partner may also be a nucleic acid generated by this invention, including a man-made gene or a man-made gene pathway. Recombination may be facilitated by or may occur at areas of homology that exist at the one (or more) artificially introduced intron(s) in the man-made gene.

The synthetic ligation reassembly method of this invention utilizes a plurality of nucleic acid building blocks, each of which preferably has two ligatable ends. The two ligatable ends on each nucleic acid building block may be two blunt ends (i.e. each having an overhang of zero nucleotides), or preferably one blunt end and one overhang, or more preferably still two overhangs.

A serviceable overhang for this purpose may be a 3' overhang or a 5' overhang. Thus, a nucleic acid building block may have a 3' overhang or alternatively a 5' overhang or alternatively two 3' overhangs or alternatively two 5' overhangs. The overall order in which the nucleic acid building blocks are assembled to form a finalized chimeric nucleic acid molecule is determined by purposeful experimental design and is not random.

According to one preferred embodiment, a nucleic acid building block is generated by chemical synthesis of two single-stranded nucleic acids (also referred to as single-stranded oligos) and contacting them so as to allow them to anneal to form a double-stranded nucleic acid building block.

A double-stranded nucleic acid building block can be of variable size. The sizes of these building blocks can be small or large depending on the choice of the experimenter. Preferred sizes for building block range from 1 base pair (not including any overhangs) to 100,000 base pairs (not including any overhangs). Other preferred size

ranges are also provided, which have lower limits of from 1 bp to 10,000 bp (including every integer value in between), and upper limits of from 2 bp to 100, 000 bp (including every integer value in between).

It is appreciated that current methods of polymerase-based amplification can be used to generate double-stranded nucleic acids of up to thousands of base pairs, if not tens of thousands of base pairs, in length with high fidelity. Chemical synthesis (e.g. phosphoramidite-based) can be used to generate nucleic acids of up to hundreds of nucleotides in length with high fidelity; however, these can be assembled, e.g. using overhangs or sticky ends, to form double-stranded nucleic acids of up to thousands of base pairs, if not tens of thousands of base pairs, in length if so desired.

A combination of methods (e.g. phosphoramidite-based chemical synthesis and PCR) can also be used according to this invention. Thus, nucleic acid building block made by different methods can also be used in combination to generate a progeny molecule of this invention.

The use of chemical synthesis to generate nucleic acid building blocks is particularly preferred in this invention & is advantageous for other reasons as well, including procedural safety and ease. No cloning or harvesting or actual handling of any biological samples is required. The design of the nucleic acid building blocks can be accomplished on paper. Accordingly, this invention teaches an advance in procedural safety in recombinant technologies.

Nonetheless, according to one preferred embodiment, a double-stranded nucleic acid building block according to this invention may also be generated by polymerase-based amplification of a polynucleotide template. In a non-limiting exemplification, as illustrated in Figure 2, a first polymerase-based amplification reaction using a first set of primers, F_2 and R_1, is used to generate a blunt-ended product (labeled Reaction 1, Product

1), which is essentially identical to Product A. A second polymerase-based amplification reaction using a second set of primers, F_1 and R_2, is used to generate a blunt-ended product (labeled Reaction 2, Product 2), which is essentially identical to Product B. These two products are mixed and allowed to melt and anneal, generating potentially useful double-stranded nucleic acid building blocks with two overhangs. In the example of Fig. 2, the product with the 3' overhangs (Product C) is selected by nuclease-based degradation of the other 3 products using a 3' acting exonuclease, such as exonuclease III. It is appreciated that a 5' acting exonuclease (e.g. red alpha) may be also be used, for example to select Product D instead. It is also appreciated that other selection means can also be used, including hybridization-based means, and that these means can incorporate a further means, such as a magnetic bead-based means, to facilitate separation of the desired product.

Many other methods exist by which a double-stranded nucleic acid building block can be generated that is serviceable for this invention; and these are known in the art and can be readily performed by the skilled artisan.

According to particularly preferred embodiment, a double-stranded nucleic acid building block that is serviceable for this invention is generated by first generating two single stranded nucleic acids and allowing them to anneal to form a double-stranded nucleic acid building block. The two strands of a double-stranded nucleic acid building block may be complementary at every nucleotide apart from any that form an overhang; thus containing no mismatches, apart from any overhang(s). According to another embodiment, the two strands of a double-stranded nucleic acid building block are complementary at fewer than every nucleotide apart from any that form an overhang. Thus, according to this embodiment, a double-stranded nucleic acid building block can be used to introduce codon degeneracy. Preferably the codon degeneracy is introduced using the site-saturation mutagenesis described herein, using one or more N,N,G/T cassettes or alternatively using one or more N,N,N cassettes.

Contained within an exemplary experimental design for achieving an ordered assembly according to this invention are:

1) The design of specific nucleic acid building blocks.

2) The design of specific ligatable ends on each nucleic acid building block.

3) The design of a particular order of assembly of the nucleic acid building blocks.

An overhang may be a 3' overhang or a 5' overhang. An overhang may also have a terminal phosphate group or alternatively may be devoid of a terminal phosphate group (having, e.g., a hydroxyl group instead). An overhang may be comprised of any number of nucleotides. Preferably an overhang is comprised of 0 nucleotides (as in a blunt end) to 10,000 nucleotides. Thus, a wide range of overhang sizes may be serviceable. Accordingly, the lower limit may be each integer from 1-200 and the upper limit may be each integer from 2-10,000. According to a particular exemplification, an overhang may consist of anywhere from 1 nucleotide to 200 nucleotides (including every integer value in between).

The final chimeric nucleic acid molecule may be generated by sequentially assembling 2 or more building blocks at a time until all the designated building blocks have been assembled. A working sample may optionally be subjected to a process for size selection or purification or other selection or enrichment process between the performance of two assembly steps. Alternatively, the final chimeric nucleic acid molecule may be generated by assembling all the designated building blocks at once in one step.

Utility

The *in vivo* recombination method of this invention can be performed blindly on a pool of unknown hybrids or alleles of a specific polynucleotide or sequence. However, it is not necessary to know the actual DNA or RNA sequence of the specific polynucleotide.

The approach of using recombination within a mixed population of genes can be useful for the generation of any useful proteins, for example, interleukin I, antibodies, tPA and growth hormone. This approach may be used to generate proteins having altered specificity or activity. The approach may also be useful for the generation of hybrid nucleic acid sequences, for example, promoter regions, introns, exons, enhancer sequences, 31 untranslated regions or 51 untranslated regions of genes. Thus this approach may be used to generate genes having increased rates of expression. This approach may also be useful in the study of repetitive DNA sequences. Finally, this approach may be useful to mutate ribozymes or aptamers.

Scaffold-like regions separating regions of diversity in proteins may be particularly suitable for the methods of this invention. The conserved scaffold determines the overall folding by self-association, while displaying relatively unrestricted loops that mediate the specific binding. Examples of such scaffolds are the immunoglobulin beta barrel, and the four-helix bundle. The methods of this invention can be used to create scaffold-like proteins with various combinations of mutated sequences for binding.

The equivalents of some standard genetic matings may also be performed by the methods of this invention. For example, a "molecular" backcross can be performed by repeated mixing of the hybrid's nucleic acid with the wild-type nucleic acid while selecting for the mutations of interest. As in traditional breeding, this approach can be used to combine phenotypes from different sources into a background of choice. It is

useful, for example, for the removal of neutral mutations that affect unselected characteristics (i.e. immunogenicity). Thus it can be useful to determine which mutations in a protein are involved in the enhanced biological activity and which are not.

5

2.11.2.4. END-SELECTION

This invention provides a method for selecting a subset of polynucleotides from a starting set of polynucleotides, which method is based on the ability to discriminate one
10 or more selectable features (or selection markers) present anywhere in a working polynucleotide, so as to allow one to perform selection for (positive selection) &/or against (negative selection) each selectable polynucleotide. In a preferred aspect, a method is provided termed end-selection, which method is based on the use of a selection marker located in part or entirely in a terminal region of a selectable polynucleotide, and
15 such a selection marker may be termed an "end-selection marker".

End-selection may be based on detection of naturally occurring sequences or on detection of sequences introduced experimentally (including by any mutagenesis procedure mentioned herein and not mentioned herein) or on both, even within the same
20 polynucleotide. An end-selection marker can be a structural selection marker or a functional selection marker or both a structural and a functional selection marker. An end-selection marker may be comprised of a polynucleotide sequence or of a polypeptide sequence or of any chemical structure or of any biological or biochemical tag, including markers that can be selected using methods based on the detection of radioactivity, of
25 enzymatic activity, of fluorescence, of any optical feature, of a magnetic property (e.g. using magnetic beads), of immunoreactivity, and of hybridization.

End-selection may be applied in combination with any method serviceable for performing mutagenesis. Such mutagenesis methods include, but are not limited to,

methods described herein (supra and infra). Such methods include, by way of non-limiting exemplification, any method that may be referred herein or by others in the art by any of the following terms: "saturation mutagenesis", "shuffling", "recombination", "re-assembly", "error-prone PCR", "assembly PCR", "sexual PCR", "crossover PCR", "oligonucleotide primer-directed mutagenesis", "recursive (&/or exponential) ensemble mutagenesis (see **Arkin** and **Youvan**, 1992)", "cassette mutagenesis", "in vivo mutagenesis", and "in vitro mutagenesis". Moreover, end-selection may be performed on molecules produced by any mutagenesis &/or amplification method (see, e.g., **Arnold**, 1993; **Caldwell** and **Joyce**, 1992; **Stemmer**, 1994; following which method it is desirable to select for (including to screen for the presence of) desirable progeny molecules.

In addition, end-selection may be applied to a polynucleotide apart from any mutagenesis method. In a preferred embodiment, end-selection, as provided herein, can be used in order to facilitate a cloning step, such as a step of ligation to another polynucleotide (including ligation to a vector). This invention thus provides for end-selection as a serviceable means to facilitate library construction, selection &/or enrichment for desirable polynucleotides, and cloning in general.

In a particularly preferred embodiment, end-selection can be based on (positive) selection for a polynucleotide; alternatively end-selection can be based on (negative) selection against a polynucleotide; and alternatively still, end-selection can be based on both (positive) selection for, and on (negative) selection against, a polynucleotide. End-selection, along with other methods of selection &/or screening, can be performed in an iterative fashion, with any combination of like or unlike selection &/or screening methods and serviceable mutagenesis methods, all of which can be performed in an iterative fashion and in any order, combination, and permutation.

It is also appreciated that, according to one embodiment of this invention, end-selection may also be used to select a polynucleotide is at least in part: circular (e.g. a

plasmid or any other circular vector or any other polynucleotide that is partly circular), &/or branched, &/or modified or substituted with any chemical group or moiety. In accord with this embodiment, a polynucleotide may be a circular molecule comprised of an intermediate or central region, which region is flanked on a 5' side by a 5' flanking region (which, for the purpose of end-selection, serves in like manner to a 5' terminal region of a non-circular polynucleotide) and on a 3' side by a 3' terminal region (which, for the purpose of end-selection, serves in like manner to a 3' terminal region of a non-circular polynucleotide). As used in this non-limiting exemplification, there may be sequence overlap between any two regions or even among all three regions.

In one non-limiting aspect of this invention, end-selection of a linear polynucleotide is performed using a general approach based on the presence of at least one end-selection marker located at or near a polynucleotide end or terminus (that can be either a 5' end or a 3' end). In one particular non-limiting exemplification, end-selection is based on selection for a specific sequence at or near a terminus such as, but not limited to, a sequence recognized by an enzyme that recognizes a polynucleotide sequence. An enzyme that recognizes and catalyzes a chemical modification of a polynucleotide is referred to herein as a polynucleotide-acting enzyme. In a preferred embodiment, serviceable polynucleotide-acting enzymes are exemplified non-exclusively by enzymes with polynucleotide-cleaving activity, enzymes with polynucleotide-methylating activity, enzymes with polynucleotide-ligating activity, and enzymes with a plurality of distinguishable enzymatic activities (including non-exclusively, e.g., both polynucleotide-cleaving activity and polynucleotide-ligating activity).

Relevant polynucleotide-acting enzymes thus also include any commercially available or non-commercially available polynucleotide endonucleases and their companion methylases including those catalogued at the website http://www.neb.com/rebase, and those mentioned in the following cited reference (**Roberts** and **Macelis**, 1996). Preferred polynucleotide endonucleases include – but are

not limited to – type II restriction enzymes (including type IIS), and include enzymes that cleave both strands of a double stranded polynucleotide (e.g. *Not* I, which cleaves both strands at 5'...GC/GGCCGC...3') and enzymes that cleave only one strand of a double stranded polynucleotide, i.e. enzymes that have polynucleotide-nicking activity, (e.g. N. *Bst*NB I, which cleaves only one strand at 5'...GAGTCNNNN/N...3'). Relevant polynucleotide-acting enzymes also include type III restriction enzymes.

It is appreciated that relevant polynucleotide-acting enzymes also include any enzymes that may be developed in the future, though currently unavailable, that are serviceable for generating a ligation compatible end, preferably a sticky end, in a polynucleotide.

In one preferred exemplification, a serviceable selection marker is a restriction site in a polynucleotide that allows a corresponding type II (or type IIS) restriction enzyme to cleave an end of the polynucleotide so as to provide a ligatable end (including a blunt end or alternatively a sticky end with at least a one base overhang) that is serviceable for a desirable ligation reaction without cleaving the polynucleotide internally in a manner that destroys a desired internal sequence in the polynucleotide. Thus it is provided that, among relevant restriction sites, those sites that do not occur internally (i.e. that do not occur apart from the termini) in a specific working polynucleotide are preferred when the use of a corresponding restriction enzyme(s) is not intended to cut the working polynucleotide internally. This allows one to perform restriction digestion reactions to completion or to near completion without incurring unwanted internal cleavage in a working polynucleotide.

According to a preferred aspect, it is thus preferable to use restriction sites that are not contained, or alternatively that are not expected to be contained, or alternatively that unlikely to be contained (e.g. when sequence information regarding a working polynucleotide is incomplete) internally in a polynucleotide to be subjected to end-

selection. In accordance with this aspect, it is appreciated that restriction sites that occur relatively infrequently are usually preferred over those that occur more frequently. On the other hand it is also appreciated that there are occasions where internal cleavage of a polypeptide is desired, e.g. to achieve recombination or other mutagenic procedures along with end-selection.

In accord with this invention, it is also appreciated that methods (e.g. mutagenesis methods) can be used to remove unwanted internal restriction sites. It is also appreciated that a partial digestion reaction (i.e. a digestion reaction that proceeds to partial completion) can be used to achieve digestion at a recognition site in a terminal region while sparing a susceptible restriction site that occurs internally in a polynucleotide and that is recognized by the same enzyme. In one aspect, partial digest are useful because it is appreciated that certain enzymes show preferential cleavage of the same recognition sequence depending on the location and environment in which the recognition sequence occurs. For example, it is appreciated that, while lambda DNA has 5 *Eco*R I sites, cleavage of the site nearest to the right terminus has been reported to occur 10 times faster than the sites in the middle of the molecule. Also, for example, it has been reported that, while *Sac* II has four sites on lambda DNA, the three clustered centrally in lambda are cleaved 50 times faster than the remaining site near the terminus (at nucleotide 40,386). Summarily, site preferences have been reported for various enzymes by many investigators (e.g., **Thomas** and **Davis**, 1975; **Forsblum** et al, 1976; **Nath** and **Azzolina**, 1981; **Brown** and **Smith**, 1977; **Gingeras** and **Brooks**, 1983; **Krüger** et al, 1988; **Conrad** and **Topal**, 1989; **Oller** et al, 1991; **Topal**, 1991; and **Pein**, 1991; to name but a few). It is appreciated that any empirical observations as well as any mechanistic understandings of site preferences by any serviceable polynucleotide-acting enzymes, whether currently available or to be procured in the future, may be serviceable in end-selection according to this invention.

It is also appreciated that protection methods can be used to selectively protect specified restriction sites (e.g. internal sites) against unwanted digestion by enzymes that would otherwise cut a working polypeptide in response to the presence of those sites; and that such protection methods include modifications such as methylations and base substitutions (e.g. U instead of T) that inhibit an unwanted enzyme activity. It is appreciated that there are limited numbers of available restriction enzymes that are rare enough (e.g. having very long recognition sequences) to create large (e.g. megabase-long) restriction fragments, and that protection approaches (e.g. by methylation) are serviceable for increasing the rarity of enzyme cleavage sites. The use of M.*Fnu* II (mCGCG) to increase the apparent rarity of *Not* I approximately twofold is but one example among many (**Qiang** et al, 1990; **Nelson** et al, 1984; **Maxam** and **Gilbert**, 1980; **Raleigh** and **Wilson**, 1986).

According to a preferred aspect of this invention, it is provided that, in general, the use of rare restriction sites is preferred. It is appreciated that, in general, the frequency of occurrence of a restriction site is determined by the number of nucleotides contained therein, as well as by the ambiguity of the base requirements contained therein. Thus, in a non-limiting exemplification, it is appreciated that, in general, a restriction site composed of, for example, 8 specific nucleotides (e.g. the *Not* I site or GC/GGCCGC, with an estimated relative occurrence of 1 in 4^8, i.e. 1 in 65,536, random 8-mers) is relatively more infrequent than one composed of, for example, 6 nucleotides (e.g. the *Sma* I site or CCC/GGG, having an estimated relative occurrence of 1 in 4^6, i.e. 1 in 4,096, random 6-mers), which in turn is relatively more infrequent than one composed of, for example, 4 nucleotides (e.g. the *Msp* I site or C/CGG, having an estimated relative occurrence of 1 in 4^4, i.e. 1 in 256, random 4-mers). Moreover, in another non-limiting exemplification, it is appreciated that, in general, a restriction site having no ambiguous (but only specific) base requirements (e.g. the *Fin* I site or GTCCC, having an estimated relative occurrence of 1 in 4^5, i.e. 1 in 1024, random 5-mers) is relatively more infrequent than one having an ambiguous W (where W = A or T) base requirement (e.g. the *Ava* II

site or G/GWCC, having an estimated relative occurrence of 1 in 4x4x2x4x4 - i.e. 1 in 512 – random 5-mers), which in turn is relatively more infrequent than one having an ambiguous N (where N = A or C or G or T) base requirement (e.g. the *Asu* I site or G/GNCC, having an estimated relative occurrence of 1 in 4x4x1x4x4, i.e. 1 in 256 – random 5-mers). These relative occurrences are considered general estimates for actual polynucleotides, because it is appreciated that specific nucleotide bases (not to mention specific nucleotide sequences) occur with dissimilar frequencies in specific polynucleotides, in specific species of organisms, and in specific groupings of organisms. For example, it is appreciated that the % G+C contents of different species of organisms are often very different and wide ranging.

The use of relatively more infrequent restriction sites as a selection marker include - in a non-limiting fashion - preferably those sites composed at least a 4 nucleotide sequence, more preferably those composed at least a 5 nucleotide sequence, more preferably still those composed at least a 6 nucleotide sequence (e.g. the *Bam*H I site or G/GATCC, the *Bgl* II site or A/GATCT, the *Pst* I site or CTGCA/G, and the *Xba* I site or T/CTAGA), more preferably still those composed at least a 7 nucleotide sequence, more preferably still those composed of an 8 nucleotide sequence nucleotide sequence (e.g. the *Asc* I site or GG/CGCGCC, the *Not* I site or GC/GGCCGC, the *Pac* I site or TTAAT/TAA, the *Pme* I site or GTTT/AAAC, the *Srf* I site or GCCC/GGGC, the *Sse*838 I site or CCTGCA/GG, and the *Swa* I site or ATTT/AAAT), more preferably still those composed of a 9 nucleotide sequence, and even more preferably still those composed of at least a 10 nucleotide sequence (e.g. the *Bsp*G I site or CG/CGCTGGAC). It is further appreciated that some restriction sites (e.g. for class IIS enzymes) are comprised of a portion of relatively high specificity (i.e. a portion containing a principal determinant of the frequency of occurrence of the restriction site) and a portion of relatively low specificity; and that a site of cleavage may or may not be contained within a portion of relatively low specificity. For example, in the *Eco*57 I site or CTGAAG(16/14), there is a

portion of relatively high specificity (i.e. the CTGAAG portion) and a portion of relatively low specificity (i.e. the N16 sequence) that contains a site of cleavage.

In another preferred embodiment of this invention, a serviceable end-selection marker is a terminal sequence that is recognized by a polynucleotide-acting enzyme that recognizes a specific polynucleotide sequence. In a preferred aspect of this invention, serviceable polynucleotide-acting enzymes also include other enzymes in addition to classic type II restriction enzymes. According to this preferred aspect of this invention, serviceable polynucleotide-acting enzymes also include gyrases, helicases, recombinases, relaxases, and any enzymes related thereto.

Among preferred examples are topoisomerases (which have been categorized by some as a subset of the gyrases) and any other enzymes that have polynucleotide-cleaving activity (including preferably polynucleotide-nicking activity) &/or polynucleotide-ligating activity. Among preferred topoisomerase enzymes are topoisomerase I enzymes, which is available from many commercial sources (Epicentre Technologies, Madison, WI; Invitrogen, Carlsbad, CA; Life Technologies, Gathesburg, MD) and conceivably even more private sources. It is appreciated that similar enzymes may be developed in the future that are serviceable for end-selection as provided herein. A particularly preferred topoisomerase I enzyme is a topoisomerase I enzyme of vaccinia virus origin, that has a specific recognition sequence (e.g. 5'...AAGGG...3') and has both polynucleotide-nicking activity and polynucleotide-ligating activity. Due to the specific nicking-activity of this enzyme (cleavage of one strand), internal recognition sites are not prone to polynucleotide destruction resulting from the nicking activity (but rather remain annealed) at a temperature that causes denaturation of a terminal site that has been nicked. Thus for use in end-selection, it is preferable that a nicking site for topoisomerase-based end-selection be no more than 100 nucleotides from a terminus, more preferably no more than 50 nucleotides from a terminus, more preferably still no more than 25 nucloetides from a terminus, even more preferably still no more than 20

nucleotides from a terminus, even more preferably still no more than 15 nucleotides from a terminus, even more preferably still no more than 10 nucleotides from a terminus, even more preferably still no more than 8 nucleotides from a terminus, even more preferably still no more than 6 nucleotides from a terminus, and even more preferably still no more than 4 nucleotides from a terminus.

In a particularly preferred exemplification that is non-limiting yet clearly illustrative, it is appreciated that when a nicking site for topoisomerase-based end-selection is 4 nucleotides from a terminus, nicking produces a single stranded oligo of 4 bases (in a terminal region) that can be denatured from its complementary strand in an end-selectable polynucleotide; this provides a sticky end (comprised of 4 bases) in a polynucleotide that is serviceable for an ensuing ligation reaction. To accomplish ligation to a cloning vector (preferably an expression vector), compatible sticky ends can be generated in a cloning vector by any means including by restriction enzyme-based means. The terminal nucleotides (comprised of 4 terminal bases in this specific example) in an end-selectable polynucleotide terminus are thus wisely chosen to provide compatibility with a sticky end generated in a cloning vector to which the polynucleotide is to be ligated.

On the other hand, internal nicking of an end-selectable polynucleotide, e.g. 500 bases from a terminus, produces a single stranded oligo of 500 bases that is not easily denatured from its complementary strand, but rather is serviceable for repair (e.g. by the same topoisomerase enzyme that produced the nick).

This invention thus provides a method - e.g. that is vaccinia topoisomerase-based &/or type II (or IIS) restriction endonuclease-based &/or type III restriction endonuclease-based &/or nicking enzyme-based (e.g. using N. *Bst*NB I) – for producing a sticky end in a working polynucleotide, which end is ligation compatible, and which end can be comprised of at least a 1 base overhang. Preferably such a sticky end is

comprised of at least a 2-base overhang, more preferably such a sticky end is comprised of at least a 3-base overhang, more preferably still such a sticky end is comprised of at least a 4-base overhang, even more preferably still such a sticky end is comprised of at least a 5-base overhang, even more preferably still such a sticky end is comprised of at least a 6-base overhang. Such a sticky end may also be comprised of at least a 7-base overhang, or at least an 8-base overhang, or at least a 9-base overhang, or at least a 10-base overhang, or at least 15-base overhang, or at least a 20-base overhang, or at least a 25-base overhang, or at least a 30-base overhang. These overhangs can be comprised of any bases, including A, C, G, or T.

It is appreciated that sticky end overhangs introduced using topoisomerase or a nicking enzyme (e.g. using N. *Bst*NB I) can be designed to be unique in a ligation environment, so as to prevent unwanted fragment reassemblies, such as self-dimerizations and other unwanted concatamerizations.

According to one aspect of this invention, a plurality of sequences (which may but do not necessarily overlap) can be introduced into a terminal region of an end-selectable polynucleotide by the use of an oligo in a polymerase-based reaction. In a relevant, but by no means limiting example, such an oligo can be used to provide a preferred 5' terminal region that is serviceable for topoisomerase I-based end-selection, which oligo is comprised of: a 1-10 base sequence that is convertible into a sticky end (preferably by a vaccinia topoisomerase I), a ribosome binding site (i.e. and "RBS", that is preferably serviceable for expression cloning), and optional linker sequence followed by an ATG start site and a template-specific sequence of 0-100 bases (to facilitate annealment to the template in the a polymerase-based reaction). Thus, according to this example, a serviceable oligo (which may be termed a forward primer) can have the sequence: 5'[terminal sequence = $(N)_{1-10}$][topoisomerase I site & RBS = AAGGGAGGAG][linker = $(N)_{1-100}$][start codon and template-specific sequence = ATG$(N)_{0-100}$]3'.

Analogously, in a relevant, but by no means limiting example, an oligo can be used to provide a preferred 3' terminal region that is serviceable for topoisomerase I-based end-selection, which oligo is comprised of: a 1-10 base sequence that is convertible into a sticky end (preferably by a vaccinia topoisomerase I), and optional linker sequence followed by a template-specific sequence of 0-100 bases (to facilitate annealment to the template in the a polymerase-based reaction). Thus, according to this example, a serviceable oligo (which may be termed a reverse primer) can have the sequence: 5'[terminal sequence = $(N)_{1-10}$][topoisomerase I site = AAGGG][linker = $(N)_{1-100}$][template-specific sequence = $(N)_{0-100}$]3'.

It is appreciated that, end-selection can be used to distinguish and separate parental template molecules (e.g. to be subjected to mutagenesis) from progeny molecules (e.g. generated by mutagenesis). For example, a first set of primers, lacking in a topoisomerase I recognition site, can be used to modify the terminal regions of the parental molecules (e.g. in polymerase-based amplification). A different second set of primers (e.g. having a topoisomerase I recognition site) can then be used to generate mutated progeny molecules (e.g. using any polynucleotide chimerization method, such as interrupted synthesis, template-switching polymerase-based amplification, or interrupted synthesis; or using saturation mutagenesis; or using any other method for introducing a topoisomerase I recognition site into a mutagenized progeny molecule as disclosed herein) from the amplified template molecules. The use of topoisomerase I-based end-selection can then facilitate, not only discernment, but selective topoisomerase I-based ligation of the desired progeny molecules.

Annealment of a second set of primers to thusly amplified parental molecules can be facilitated by including sequences in a first set of primers (i.e. primers used for amplifying a set parental molecules) that are similar to a toposiomerase I recognition site, yet different enough to prevent functional toposiomerase I enzyme recognition. For example, sequences that diverge from the AAGGG site by anywhere from 1 base to all 5 bases can be incorporated into a first set of primers (to be used for amplifying the parental templates prior to subjection

to mutagenesis). In a specific, but non-limiting aspect, it is thus provided that a parental molecule can be amplified using the following exemplary – but by no means limiting – set of forward and reverse primers:

Forward Primer: 5' CTAGAAGAGAGGAGAAAACCATG$(N)_{10-100}$ 3', and
Reverse Primer: 5' GATCAAAGGCGCGCCTGCAGG$(N)_{10-100}$ 3'

According to this specific example of a first set of primers, $(N)_{10-100}$ represents preferably a 10 to 100 nucleotide-long template-specific sequence, more preferably a 10 to 50 nucleotide-long template-specific sequence, more preferably still a 10 to 30 nucleotide-long template-specific sequence, and even more preferably still a 15 to 25 nucleotide-long template-specific sequence.

According to a specific, but non-limiting aspect, it is thus provided that, after this amplification (using a disclosed first set of primers lacking in a true topoisomerase I recognition site), amplified parental molecules can then be subjected to mutagenesis using one or more sets of forward and reverse primers that do have a true topoisomerase I recognition site. In a specific, but non-limiting aspect, it is thus provided that a parental molecule can be used as templates for the generation of a mutagenized progeny molecule using the following exemplary – but by no means limiting – second set of forward and reverse primers:

Forward Primer: 5' CTAGAAGGGAGGAGAAAACCATG 3'
Reverse Primer: 5' GATCAAAGGCGCGCCTGCAGG 3' (contains *Asc* I recognition sequence)

It is appreciated that any number of different primers sets not specifically mentioned can be used as first, second, or subsequent sets of primers for end-selection consistent with this invention. Notice that type II restriction enzyme sites can be incorporated (e.g. an *Asc* I site in the above example). It is provided that, in addition to the other sequences mentioned,

the experimentalist can incorporate one or more N,N,G/T triplets into a serviceable primer in order to subject a working polynucleotide to saturation mutagenesis. Summarily, use of a second and/or subsequent set of primers can achieve dual goals of introducing a topoisomerase I site and of generating mutations in a progeny polynucleotide.

Thus, according to one use provided, a serviceable end-selection marker is an enzyme recognition site that allows an enzyme to cleave (including nick) a polynucleotide at a specified site, to produce a ligation-compatible end upon denaturation of a generated single stranded oligo. Ligation of the produced polynucleotide end can then be accomplished by the same enzyme (e.g. in the case of vaccinia virus topoisomerase I), or alternatively with the use of a different enzyme. According to one aspect of this invention, any serviceable end-selection markers, whether like (e.g. two vaccinia virus topoisomerase I recognition sites) or unlike (e.g. a class II restriction enzyme recognition site and a vaccinia virus topoisomerase I recognition site) can be used in combination to select a polynucleotide. Each selectable polynucleotide can thus have one or more end-selection markers, and they can be like or unlike end-selection markers. In a particular aspect, a plurality of end-selection markers can be located on one end of a polynucleotide and can have overlapping sequences with each other.

It is important to emphasize that any number of enzymes, whether currently in existence or to be developed, can be serviceable in end-selection according to this invention. For example, in a particular aspect of this invention, a nicking enzyme (e.g. N. *Bst*NB I, which cleaves only one strand at 5'...GAGTCNNNN/N...3') can be used in conjunction with a source of polynucleotide-ligating activity in order to achieve end-selection. According to this embodiment, a recognition site for N. *Bst*NB I – instead of a recognition site for topoisomerase I – should be incorporated into an end-selectable polynucleotide (whether end-selection is used for selection of a mutagenized progeny molecule or whether end-selection is used apart from any mutagenesis procedure).

It is appreciated that the instantly disclosed end-selection approach using topoisomerase-based nicking and ligation has several advantages over previously available selection methods. In sum, this approach allows one to achieve direction cloning (including expression cloning). Specifically, this approach can be used for the achievement of: direct ligation (i.e. without subjection to a classic restriction-purification-ligation reaction, that is susceptible to a multitude of potential problems from an initial restriction reaction to a ligation reaction dependent on the use of T4 DNA ligase); separation of progeny molecules from original template molecules (e.g. original template molecules lack topoisomerase I sites that not introduced until after mutagenesis), obviation of the need for size separation steps (e.g. by gel chromatography or by other electrophoretic means or by the use of size-exclusion membranes), preservation of internal sequences (even when topoisomerase I sites are present), obviation of concerns about unsuccessful ligation reactions (e.g. dependent on the use of T4 DNA ligase, particularly in the presence of unwanted residual restriction enzyme activity), and facilitated expression cloning (including obviation of frame shift concerns). Concerns about unwanted restriction enzyme-based cleavages – especially at internal restriction sites (or even at often unpredictable sites of unwanted star activity) in a working polynucleotide – that are potential sites of destruction of a working polynucleotide can also be obviated by the instantly disclosed end-selection approach using topoisomerase-based nicking and ligation.

2.11.3. ADDITIONAL SCREENING METHODS

Peptide Display Methods

The present method can be used to shuffle, by *in vitro* and/or *in vivo* recombination by any of the disclosed methods, and in any combination, polynucleotide sequences selected by peptide display methods, wherein an associated polynucleotide encodes a displayed peptide which is screened for a phenotype (e.g., for affinity for a predetermined receptor (ligand).

An increasingly important aspect of bio-pharmaceutical drug development and molecular biology is the identification of peptide structures, including the primary amino acid sequences, of peptides or peptidomimetics that interact with biological macromolecules. one method of identifying peptides that possess a desired structure or functional property, such as binding to a predetermined biological macromolecule (e.g., a receptor), involves the screening of a large library or peptides for individual library members which possess the desired structure or functional property conferred by the amino acid sequence of the peptide.

In addition to direct chemical synthesis methods for generating peptide libraries, several recombinant DNA methods also have been reported. One type involves the display of a peptide sequence, antibody, or other protein on the surface of a bacteriophage particle or cell. Generally, in these methods each bacteriophage particle or cell serves as an individual library member displaying a single species of displayed peptide in addition to the natural bacteriophage or cell protein sequences. Each bacteriophage or cell contains the nucleotide sequence information encoding the particular displayed peptide sequence; thus, the displayed peptide sequence can be ascertained by nucleotide sequence determination of an isolated library member.

A well-known peptide display method involves the presentation of a peptide sequence on the surface of a filamentous bacteriophage, typically as a fusion with a bacteriophage coat protein. The bacteriophage library can be incubated with an immobilized, predetermined macromolecule or small molecule (e.g., a receptor) so that bacteriophage particles which present a peptide sequence that binds to the immobilized macromolecule can be differentially partitioned from those that do not present peptide sequences that bind to the predetermined macromolecule. The bacteriophage particles (i.e., library members) which are bound to the immobilized macromolecule are then recovered and replicated to amplify the selected bacteriophage sub-population for a

subsequent round of affinity enrichment and phage replication. After several rounds of affinity enrichment and phage replication, the bacteriophage library members that are thus selected are isolated and the nucleotide sequence encoding the displayed peptide sequence is determined, thereby identifying the sequence(s) of peptides that bind to the predetermined macromolecule (e.g., receptor). Such methods are further described in PCT patent publications **WO 91/17271**, **WO 91/18980**, **WO 91/19818** and **WO 93/08278**.

The latter PCT publication describes a recombinant DNA method for the display of peptide ligands that involves the production of a library of fusion proteins with each fusion protein composed of a first polypeptide portion, typically comprising a variable sequence, that is available for potential binding to a predetermined macromolecule, and a second polypeptide portion that binds to DNA, such as the DNA vector encoding the individual fusion protein. When transformed host cells are cultured under conditions that allow for expression of the fusion protein, the fusion protein binds to the DNA vector encoding it. Upon lysis of the host cell, the fusion protein/vector DNA complexes can be screened against a predetermined macromolecule in much the same way as bacteriophage particles are screened in the phage-based display system, with the replication and sequencing of the DNA vectors in the selected fusion protein/vector DNA complexes serving as the basis for identification of the selected library peptide sequence(s).

Other systems for generating libraries of peptides and like polymers have aspects of both the recombinant and *in vitro* chemical synthesis methods. In these hybrid methods, cell-free enzymatic machinery is employed to accomplish the *in vitro* synthesis of the library members (i.e., peptides or polynucleotides). In one type of method, RNA molecules with the ability to bind a predetermined protein or a predetermined dye molecule were selected by alternate rounds of selection and PCR amplification (**Tuerk and Gold**, 1990; **Ellington and Szostak**, 1990). A similar technique was used to identify DNA sequences which bind a predetermined human transcription factor

(**Thiesen and Bach**, 1990; **Beaudry and Joyce**, 1992; PCT patent publications **WO 92/05258** and **WO 92/14843**). In a similar fashion, the technique of *in vitro* translation has been used to synthesize proteins of interest and has been proposed as a method for generating large libraries of peptides. These methods which rely upon *in vitro* translation, generally comprising stabilized polysome complexes, are described further in PCT patent publications **WO 88/08453, WO 90/05785, WO 90/07003, WO 91/02076, WO 91/05058**, and **WO 92/02536**. Applicants have described methods in which library members comprise a fusion protein having a first polypeptide portion with DNA binding activity and a second polypeptide portion having the library member unique peptide sequence; such methods are suitable for use in cell-free *in vitro* selection formats, among others.

The displayed peptide sequences can be of varying lengths, typically from 3-5000 amino acids long or longer, frequently from 5-100 amino acids long, and often from about 8-15 amino acids long. A library can comprise library members having varying lengths of displayed peptide sequence, or may comprise library members having a fixed length of displayed peptide sequence. Portions or all of the displayed peptide sequence(s) can be random, pseudorandom, defined set kernal, fixed, or the like. The present display methods include methods for *in vitro* and *in vivo* display of single-chain antibodies, such as nascent scFv on polysomes or scfv displayed on phage, which enable large-scale screening of scfv libraries having broad diversity of variable region sequences and binding specificities.

The present invention also provides random, pseudorandom, and defined sequence framework peptide libraries and methods for generating and screening those libraries to identify useful compounds (e.g., peptides, including single-chain antibodies) that bind to receptor molecules or epitopes of interest or gene products that modify peptides or RNA in a desired fashion. The random, pseudorandom, and defined sequence framework peptides are produced from libraries of peptide library members that comprise

displayed peptides or displayed single-chain antibodies attached to a polynucleotide template from which the displayed peptide was synthesized. The mode of attachment may vary according to the specific embodiment of the invention selected, and can include encapsulation in a phage particle or incorporation in a cell.

A method of affinity enrichment allows a very large library of peptides and single-chain antibodies to be screened and the polynucleotide sequence encoding the desired peptide(s) or single-chain antibodies to be selected. The polynucleotide can then be isolated and shuffled to recombine combinatorially the amino acid sequence of the selected peptide(s) (or predetermined portions thereof) or single-chain antibodies (or just VHI, VLI or CDR portions thereof). Using these methods, one can identify a peptide or single-chain antibody as having a desired binding affinity for a molecule and can exploit the process of shuffling to converge rapidly to a desired high-affinity peptide or scfv. The peptide or antibody can then be synthesized in bulk by conventional means for any suitable use (e.g., as a therapeutic or diagnostic agent).

A significant advantage of the present invention is that no prior information regarding an expected ligand structure is required to isolate peptide ligands or antibodies of interest. The peptide identified can have biological activity, which is meant to include at least specific binding affinity for a selected receptor molecule and, in some instances, will further include the ability to block the binding of other compounds, to stimulate or inhibit metabolic pathways, to act as a signal or messenger, to stimulate or inhibit cellular activity, and the like.

The present invention also provides a method for shuffling a pool of polynucleotide sequences selected by affinity screening a library of polysomes displaying nascent peptides (including single-chain antibodies) for library members which bind to a predetermined receptor (e.g., a mammalian proteinaceous receptor such as, for example, a peptidergic hormone receptor, a cell surface receptor, an intracellular protein which binds

to other protein(s) to form intracellular protein complexes such as hetero-dimers and the like) or epitope (e.g., an immobilized protein, glycoprotein, oligosaccharide, and the like).

Polynucleotide sequences selected in a first selection round (typically by affinity selection for binding to a receptor (e.g., a ligand)) by any of these methods are pooled and the pool(s) is/are shuffled by *in vitro* and/or *in vivo* recombination to produce a shuffled pool comprising a population of recombined selected polynucleotide sequences. The recombined selected polynucleotide sequences are subjected to at least one subsequent selection round. The polynucleotide sequences selected in the subsequent selection round(s) can be used directly, sequenced, and/or subjected to one or more additional rounds of shuffling and subsequent selection. Selected sequences can also be back-crossed with polynucleotide sequences encoding neutral sequences (i.e., having insubstantial functional effect on binding), such as for example by back-crossing with a wild-type or naturally-occurring sequence substantially identical to a selected sequence to produce native-like functional peptides, which may be less immunogenic. Generally, during back-crossing subsequent selection is applied to retain the property of binding to the predetermined receptor (ligand).

Prior to or concomitant with the shuffling of selected sequences, the sequences can be mutagenized. In one embodiment, selected library members are cloned in a prokaryotic vector (e.g., plasmid, phagemid, or bacteriophage) wherein a collection of individual colonies (or plaques) representing discrete library members are produced. Individual selected library members can then be manipulated (e.g., by site-directed mutagenesis, cassette mutagenesis, chemical mutagenesis, PCR mutagenesis, and the like) to generate a collection of library members representing a kernal of sequence diversity based on the sequence of the selected library member. The sequence of an individual selected library member or pool can be manipulated to incorporate random mutation, pseudorandom mutation, defined kernal mutation (i.e., comprising variant and

invariant residue positions and/or comprising variant residue positions which can comprise a residue selected from a defined subset of amino acid residues), codon-based mutation, and the like, either segmentally or over the entire length of the individual selected library member sequence. The mutagenized selected library members are then shuffled by *in vitro* and/or *in vivo* recombinatorial shuffling as disclosed herein.

The invention also provides peptide libraries comprising a plurality of individual library members of the invention, wherein (1) each individual library member of said plurality comprises a sequence produced by shuffling of a pool of selected sequences, and (2) each individual library member comprises a variable peptide segment sequence or single-chain antibody segment sequence which is distinct from the variable peptide segment sequences or single-chain antibody sequences of other individual library members in said plurality (although some library members may be present in more than one copy per library due to uneven amplification, stochastic probability, or the like).

The invention also provides a product-by-process, wherein selected polynucleotide sequences having (or encoding a peptide having) a predetermined binding specificity are formed by the process of: (1) screening a displayed peptide or displayed single-chain antibody library against a predetermined receptor (e.g., ligand) or epitope (e.g., antigen macromolecule) and identifying and/or enriching library members which bind to the predetermined receptor or epitope to produce a pool of selected library members, (2) shuffling by recombination the selected library members (or amplified or cloned copies thereof) which binds the predetermined epitope and has been thereby isolated and/or enriched from the library to generate a shuffled library, and (3) screening the shuffled library against the predetermined receptor (e.g., ligand) or epitope (e.g., antigen macromolecule) and identifying and/or enriching shuffled library members which bind to the predetermined receptor or epitope to produce a pool of selected shuffled library members.

Antibody Display and Screening Methods

The present method can be used to shuffle, by *in vitro* and/or *in vivo* recombination by any of the disclosed methods, and in any combination, polynucleotide sequences selected by antibody display methods, wherein an associated polynucleotide encodes a displayed antibody which is screened for a phenotype (e.g., for affinity for binding a predetermined antigen (ligand).

Various molecular genetic approaches have been devised to capture the vast immunological repertoire represented by the extremely large number of distinct variable regions which can be present in immunoglobulin chains. The naturally-occurring germ line immunoglobulin heavy chain locus is composed of separate tandem arrays of variable segment genes located upstream of a tandem array of diversity segment genes, which are themselves located upstream of a tandem array of joining (i) region genes, which are located upstream of the constant region genes. During B lymphocyte development, V-D-J rearrangement occurs wherein a heavy chain variable region gene (VH) is formed by rearrangement to form a fused D segment followed by rearrangement with a V segment to form a V-D-J joined product gene which, if productively rearranged, encodes a functional variable region (VH) of a heavy chain. Similarly, light chain loci rearrange one of several V segments with one of several J segments to form a gene encoding the variable region (VL) of a light chain.

The vast repertoire of variable regions possible in immunoglobulins derives in part from the numerous combinatorial possibilities of joining V and i segments (and, in the case of heavy chain loci, D segments) during rearrangement in B cell development. Additional sequence diversity in the heavy chain variable regions arises from non-uniform rearrangements of the D segments during V-D-J joining and from N region addition. Further, antigen-selection of specific B cell clones selects for higher affinity variants having non-germline mutations in one or both of the heavy and light chain

variable regions; a phenomenon referred to as "affinity maturation" or "affinity sharpening". Typically, these "affinity sharpening" mutations cluster in specific areas of the variable region, most commonly in the complementarity-determining regions (CDRs).

In order to overcome many of the limitations in producing and identifying high-affinity immunoglobulins through antigen-stimulated ß cell development (i.e., immunization), various prokaryotic expression systems have been developed that can be manipulated to produce combinatorial antibody libraries which may be screened for high-affinity antibodies to specific antigens. Recent advances in the expression of antibodies in Escherichia coli and bacteriophage systems (see "alternative peptide display methods", *infra*) have raised the possibility that virtually any specificity can be obtained by either cloning antibody genes from characterized hybridomas or by de novo selection using antibody gene libraries (e.g., from Ig cDNA).

Combinatorial libraries of antibodies have been generated in bacteriophage lambda expression systems which may be screened as bacteriophage plaques or as colonies of lysogens (**Huse** et al, 1989); **Caton** and **Koprowski**, 1990; **Mullinax** et al, 1990; **Persson** et al, 1991). Various embodiments of bacteriophage antibody display libraries and lambda phage expression libraries have been described (**Kang** et al, 1991; **Clackson** et al, 1991; **McCafferty** et al, 1990; **Burton** et al, 1991; **Hoogenboom** et al, 1991; **Chang** et al, 1991; **Breitling** et al, 1991; **Marks** et al, 1991, p. 581; **Barbas** et al, 1992; **Hawkins** and **Winter**, 1992; **Marks** et al, 1992, p. 779; **Marks** et al, 1992, p. 16007; and **Lowman** et al, 1991; **Lerner** et al, 1992; all incorporated herein by reference). Typically, a bacteriophage antibody display library is screened with a receptor (e.g., polypeptide, carbohydrate, glycoprotein, nucleic acid) that is immobilized (e.g., by covalent linkage to a chromatography resin to enrich for reactive phage by affinity chromatography) and/or labeled (e.g., to screen plaque or colony lifts).

One particularly advantageous approach has been the use of so-called single-chain fragment variable (scfv) libraries (**Marks** et al, 1992, p. 779; **Winter** and **Milstein**, 1991; **Clackson** et al, 1991; **Marks** et al, 1991, p. 581; **Chaudhary** et al, 1990; **Chiswell** et al, 1992; **McCafferty** et al, 1990; and **Huston** et al, 1988). Various embodiments of scfv libraries displayed on bacteriophage coat proteins have been described.

Beginning in 1988, single-chain analogues of Fv fragments and their fusion proteins have been reliably generated by antibody engineering methods. The first step generally involves obtaining the genes encoding VH and VL domains with desired binding properties; these V genes may be isolated from a specific hybridoma cell line, selected from a combinatorial V-gene library, or made by V gene synthesis. The single-chain Fv is formed by connecting the component V genes with an oligonucleotide that encodes an appropriately designed linker peptide, such as (Gly-Gly-Gly-Gly-Ser)3 or equivalent linker peptide(s). The linker bridges the C-terminus of the first V region and N-terminus of the second, ordered as either VH-linker-VL or VL-linker-VH' In principle, the scfv binding site can faithfully replicate both the affinity and specificity of its parent antibody combining site.

Thus, scfv fragments are comprised of VH and VL domains linked into a single polypeptide chain by a flexible linker peptide. After the scfv genes are assembled, they are cloned into a phagemid and expressed at the tip of the M13 phage (or similar filamentous bacteriophage) as fusion proteins with the bacteriophage PIII (gene 3) coat protein. Enriching for phage expressing an antibody of interest is accomplished by panning the recombinant phage displaying a population scfv for binding to a predetermined epitope (e.g., target antigen, receptor).

The linked polynucleotide of a library member provides the basis for replication of the library member after a screening or selection procedure, and also provides the basis for the determination, by nucleotide sequencing, of the identity of the displayed peptide

sequence or VH and VL amino acid sequence. The displayed peptide (s) or single-chain antibody (e. g., scfv) and/or its VH and VL domains or their CDRs can be cloned and expressed in a suitable expression system. Often polynucleotides encoding the isolated VH and VL domains will be ligated to polynucleotides encoding constant regions (CH and CL) to form polynucleotides encoding complete antibodies (e.g., chimeric or fully-human), antibody fragments, and the like. Often polynucleotides encoding the isolated CDRs will be grafted into polynucleotides encoding a suitable variable region framework (and optionally constant regions) to form polynucleotides encoding complete antibodies (e.g., humanized or fully-human), antibody fragments, and the like. Antibodies can be used to isolate preparative quantities of the antigen by immunoaffinity chromatography. Various other uses of such antibodies are to diagnose and/or stage disease (e.g., neoplasia) and for therapeutic application to treat disease, such as for example: neoplasia, autoimmune disease, AIDS, cardiovascular disease, infections, and the like.

Various methods have been reported for increasing the combinatorial diversity of a scfv library to broaden the repertoire of binding species (idiotype spectrum) The use of PCR has permitted the variable regions to be rapidly cloned either from a specific hybridoma source or as a gene library from non-immunized cells, affording combinatorial diversity in the assortment of VH and VL cassettes which can be combined. Furthermore, the VH and VL cassettes can themselves be diversified, such as by random, pseudorandom, or directed mutagenesis. Typically, VH and VL cassettes are diversified in or near the complementarity-determining regions (CDRS), often the third CDR, CDR3. Enzymatic inverse PCR mutagenesis has been shown to be a simple and reliable method for constructing relatively large libraries of scfv site-directed hybrids (**Stemmer** et al, 1993), as has error-prone PCR and chemical mutagenesis (**Deng** et al, 1994). Riechmann (**Riechmann** et al, 1993) showed semi-rational design of an antibody scfv fragment using site-directed randomization by degenerate oligonucleotide PCR and subsequent phage display of the resultant scfv hybrids. Barbas (**Barbas** et al, 1992) attempted to circumvent

the problem of limited repertoire sizes resulting from using biased variable region sequences by randomizing the sequence in a synthetic CDR region of a human tetanus toxoid-binding Fab.

CDR randomization has the potential to create approximately 1 x 10^{20} CDRs for the heavy chain CDR3 alone, and a roughly similar number of variants of the heavy chain CDR1 and CDR2, and light chain CDR1-3 variants. Taken individually or together, the combination possibilities of CDR randomization of heavy and/or light chains requires generating a prohibitive number of bacteriophage clones to produce a clone library representing all possible combinations, the vast majority of which will be non-binding. Generation of such large numbers of primary transformants is not feasible with current transformation technology and bacteriophage display systems. For example, Barbas (**Barbas** et al, 1992) only generated 5 x 10^{7} transformants, which represents only a tiny fraction of the potential diversity of a library of thoroughly randomized CDRS.

Despite these substantial limitations, bacteriophage. display of scfv have already yielded a variety of useful antibodies and antibody fusion proteins. A bispecific single chain antibody has been shown to mediate efficient tumor cell lysis (**Gruber** et al, 1994). Intracellular expression of an anti-Rev scfv has been shown to inhibit HIV-1 virus replication *in vitro* (**Duan** et al, 1994), and intracellular expression of an anti-p2lrar, scfv has been shown to inhibit meiotic maturation of Xenopus oocytes (**Biocca** et al, 1993). Recombinant scfv which can be used to diagnose HIV infection have also been reported, demonstrating the diagnostic utility of scfv (**Lilley** et al, 1994). Fusion proteins wherein an scFv is linked to a second polypeptide, such as a toxin or fibrinolytic activator protein, have also been reported (**Holvost** et al, 1992; **Nicholls** et al, 1993).

If it were possible to generate scfv libraries having broader antibody diversity and overcoming many of the limitations of conventional CDR mutagenesis and randomization methods which can cover only a very tiny fraction of the potential

sequence combinations, the number and quality of scfv antibodies suitable for therapeutic and diagnostic use could be vastly improved. To address this, the *in vitro* and *in vivo* shuffling methods of the invention are used to recombine CDRs which have been obtained (typically via PCR amplification or cloning) from nucleic acids obtained from selected displayed antibodies. Such displayed antibodies can be displayed on cells, on bacteriophage particles, on polysomes, or any suitable antibody display system wherein the antibody is associated with its encoding nucleic acid(s). In a variation, the CDRs are initially obtained from mRNA (or cDNA) from antibody-producing cells (e.g., plasma cells/splenocytes from an immunized wild-type mouse, a human, or a transgenic mouse capable of making a human antibody as in **WO 92/03918**, **WO 93/12227**, and **WO 94/25585**), including hybridomas derived therefrom.

Polynucleotide sequences selected in a first selection round (typically by affinity selection for displayed antibody binding to an antigen (e.g., a ligand) by any of these methods are pooled and the pool(s) is/are shuffled by *in vitro* and/or *in vivo* recombination, especially shuffling of CDRs (typically shuffling heavy chain CDRs with other heavy chain CDRs and light chain CDRs with other light chain CDRs) to produce a shuffled pool comprising a population of recombined selected polynucleotide sequences. The recombined selected polynucleotide sequences are expressed in a selection format as a displayed antibody and subjected to at least one subsequent selection round. The polynucleotide sequences selected in the subsequent selection round(s) can be used directly, sequenced, and/or subjected to one or more additional rounds of shuffling and subsequent selection until an antibody of the desired binding affinity is obtained. Selected sequences can also be back-crossed with polynucleotide sequences encoding neutral antibody framework sequences (i.e., having insubstantial functional effect on antigen binding), such as for example by back-crossing with a human variable region framework to produce human-like sequence antibodies. Generally, during back-crossing subsequent selection is applied to retain the property of binding to the predetermined antigen.

Alternatively, or in combination with the noted variations, the valency of the target epitope may be varied to control the average binding affinity of selected scfv library members. The target epitope can be bound to a surface or substrate at varying densities, such as by including a competitor epitope, by dilution, or by other method known to those in the art. A high density (valency) of predetermined epitope can be used to enrich for scfv library members which have relatively low affinity, whereas a low density (valency) can preferentially enrich for higher affinity scfv library members.

For generating diverse variable segments, a collection of synthetic oligonucleotides encoding random, pseudorandom, or a defined sequence kernal set of peptide sequences can be inserted by ligation into a predetermined site (e.g., a CDR). Similarly, the sequence diversity of one or more CDRs of the single-chain antibody cassette(s) can be expanded by mutating the CDR(s) with site-directed mutagenesis, CDR-replacement, and the like. The resultant DNA molecules can be propagated in a host for cloning and amplification prior to shuffling, or can be used directly (i.e., may avoid loss of diversity which may occur upon propagation in a host cell) and the selected library members subsequently shuffled.

Displayed peptide/polynucleotide complexes (library members) which encode a variable segment peptide sequence of interest or a single-chain antibody of interest are selected from the library by an affinity enrichment technique. This is accomplished by means of a immobilized macromolecule or epitope specific for the peptide sequence of interest, such as a receptor, other macromolecule, or other epitope species. Repeating the affinity selection procedure provides an enrichment of library members encoding the desired sequences, which may then be isolated for pooling and shuffling, for sequencing, and/or for further propagation and affinity enrichment.

The library members without the desired specificity are removed by washing. The degree and stringency of washing required will be determined for each peptide sequence or single-chain antibody of interest and the immobilized predetermined macromolecule or epitope. A certain degree of control can be exerted over the binding characteristics of the nascent peptide/DNA complexes recovered by adjusting the conditions of the binding incubation and the subsequent washing. The temperature, pH, ionic strength, divalent cations concentration, and the volume and duration of the washing will select for nascent peptide/DNA complexes within particular ranges of affinity for the immobilized macromolecule. Selection based on slow dissociation rate, which is usually predictive of high affinity, is often the most practical route. This may be done either by continued incubation in the presence of a saturating amount of free predetermined macromolecule, or by increasing the volume, number, and length of the washes. In each case, the rebinding of dissociated nascent peptide/DNA or peptide/RNA complex is prevented, and with increasing time, nascent peptide/DNA or peptide/RNA complexes of higher and higher affinity are recovered.

Additional modifications of the binding and washing procedures may be applied to find peptides with special characteristics. The affinities of some peptides are dependent on ionic strength or cation concentration. This is a useful characteristic for peptides that will be used in affinity purification of various proteins when gentle conditions for removing the protein from the peptides are required.

One variation involves the use of multiple binding targets (multiple epitope species, multiple receptor species), such that a scfv library can be simultaneously screened for a multiplicity of scfv which have different binding specificities. Given that the size of a scfv library often limits the diversity of potential scfv sequences, it is typically desirable to us scfv libraries of as large a size as possible. The time and economic considerations of generating a number of very large polysome scFv-display libraries can become prohibitive. To avoid this substantial problem, multiple

predetermined epitope species (receptor species) can be concomitantly screened in a single library, or sequential screening against a number of epitope species can be used. In one variation, multiple target epitope species, each encoded on a separate bead (or subset of beads), can be mixed and incubated with a polysome-display scfv library under suitable binding conditions. The collection of beads, comprising multiple epitope species, can then be used to isolate, by affinity selection, scfv library members. Generally, subsequent affinity screening rounds can include the same mixture of beads, subsets thereof, or beads containing only one or two individual epitope species. This approach affords efficient screening, and is compatible with laboratory automation, batch processing, and high throughput screening methods.

A variety of techniques can be used in the present invention to diversify a peptide library or single-chain antibody library, or to diversify, prior to or concomitant with shuffling, around variable segment peptides found in early rounds of panning to have sufficient binding activity to the predetermined macromolecule or epitope. In one approach, the positive selected peptide/polynucleotide complexes (those identified in an early round of affinity enrichment) are sequenced to determine the identity of the active peptides. Oligonucleotides are then synthesized based on these active peptide sequences, employing a low level of all bases incorporated at each step to produce slight variations of the primary oligonucleotide sequences. This mixture of (slightly) degenerate oligonucleotides is then cloned into the variable segment sequences at the appropriate locations. This method produces systematic, controlled variations of the starting peptide sequences, which can then be shuffled. It requires, however, that individual positive nascent peptide/polynucleotide complexes be sequenced before mutagenesis, and thus is useful for expanding the diversity of small numbers of recovered complexes and selecting variants having higher binding affinity and/or higher binding specificity. In a variation, mutagenic PCR amplification of positive selected peptide/polynucleotide complexes (especially of the variable region sequences, the amplification products of which are shuffled *in vitro* and/or *in vivo* and one or more additional rounds of screening is done

prior to sequencing. The same general approach can be employed with single-chain antibodies in order to expand the diversity and enhance the binding affinity/specificity, typically by diversifying CDRs or adjacent framework regions prior to or concomitant with shuffling. If desired, shuffling reactions can be spiked with mutagenic oligonucleotides capable of *in vitro* recombination with the selected library members can be included. Thus, mixtures of synthetic oligonucleotides and PCR produced polynucleotides (synthesized by error-prone or high-fidelity methods) can be added to the *in vitro* shuffling mix and be incorporated into resulting shuffled library members (shufflants).

The present invention of shuffling enables the generation of a vast library of CDR-variant single-chain antibodies. One way to generate such antibodies is to insert synthetic CDRs into the single-chain antibody and/or CDR randomization prior to or concomitant with shuffling. The sequences of the synthetic CDR cassettes are selected by referring to known sequence data of human CDR and are selected in the discretion of the practitioner according to the following guidelines: synthetic CDRs will have at least 40 percent positional sequence identity to known CDR sequences, and preferably will have at least 50 to 70 percent positional sequence identity to known CDR sequences. For example, a collection of synthetic CDR sequences can be generated by synthesizing a collection of oligonucleotide sequences on the basis of naturally-occurring human CDR sequences listed in Kabat (**Kabat** et al, 1991); the pool (s) of synthetic CDR sequences are calculated to encode CDR peptide sequences having at least 40 percent sequence identity to at least one known naturally-occurring human CDR sequence. Alternatively, a collection of naturally-occurring CDR sequences may be compared to generate consensus sequences so that amino acids used at a residue position frequently (i.e., in at least 5 percent of known CDR sequences) are incorporated into the synthetic CDRs at the corresponding position(s). Typically, several (e.g., 3 to about 50) known CDR sequences are compared and observed natural sequence variations between the known CDRs are tabulated, and a collection of oligonucleotides encoding CDR peptide sequences

encompassing all or most permutations of the observed natural sequence variations is synthesized. For example but not for limitation, if a collection of human VH CDR sequences have carboxy-terminal amino acids which are either Tyr, Val, Phe, or Asp, then the pool(s) of synthetic CDR oligonucleotide sequences are designed to allow the carboxy-terminal CDR residue to be any of these amino acids. In some embodiments, residues other than those which naturally-occur at a residue position in the collection of CDR sequences are incorporated: conservative amino acid substitutions are frequently incorporated and up to 5 residue positions may be varied to incorporate non-conservative amino acid substitutions as compared to known naturally-occurring CDR sequences. Such CDR sequences can be used in primary library members (prior to first round screening) and/or can be used to spike *in vitro* shuffling reactions of selected library member sequences. Construction of such pools of defined and/or degenerate sequences will be readily accomplished by those of ordinary skill in the art.

The collection of synthetic CDR sequences comprises at least one member that is not known to be a naturally-occurring CDR sequence. It is within the discretion of the practitioner to include or not include a portion of random or pseudorandom sequence corresponding to N region addition in the heavy chain CDR; the N region sequence ranges from 1 nucleotide to about 4 nucleotides occurring at V-D and D-J junctions. A collection of synthetic heavy chain CDR sequences comprises at least about 100 unique CDR sequences, typically at least about 1,000 unique CDR sequences, preferably at least about 10,000 unique CDR sequences, frequently more than 50,000 unique CDR sequences; however, usually not more than about 1 x 10 6 unique CDR sequences are included in the collection, although occasionally 1 x 107 to 1 *X* 108 unique CDR sequences are present, especially if conservative amino acid substitutions are permitted at positions where the conservative amino acid substituent is not present or is rare (i.e., less than 0.1 percent) in that position in naturally-occurring human CDRS. In general, the number of unique CDR sequences included in a library should not exceed the expected number of primary transformants in the library by more than a factor of 10. Such

single-chain antibodies generally bind of about at least 1 x 10 m-, preferably with an affinity of about at least 5 x 10^7 M-1, more preferably with an affinity of at least 1 x 10^8 M-1 to 1 x 10^9 M-1 or more, sometimes up to 1 x 10^{10} M-1 or more. Frequently, the predetermined antigen is a human protein, such as for example a human cell surface antigen (e. g., CD4, CD8, IL-2 receptor, EGF receptor, PDGF receptor), other human biological macromolecule (e.g., thrombomodulin, protein C, carbohydrate antigen, sialyl Lewis antigen, Lselectin), or nonhuman disease associated macromolecule (e.g., bacterial LPS, virion capsid protein or envelope glycoprotein) and the like.

High affinity single-chain antibodies of the desired specificity can be engineered and expressed in a variety of systems. For example, scfv have been produced in plants (**Firek** et al, 1993) and can be readily made in prokaryotic systems (**Owens** and **Young**, 1994; **Johnson** and **Bird**, 1991). Furthermore, the single-chain antibodies can be used as a basis for constructing whole antibodies or various fragments thereof (**Kettleborough** et al, 1994). The variable region encoding sequence may be isolated (e.g., by PCR amplification or subcloning) and spliced to a sequence encoding a desired human constant region to encode a human sequence antibody more suitable for human therapeutic uses where immunogenicity is preferably minimized. The polynucleotide(s) having the resultant fully human encoding sequence(s) can be expressed in a host cell (e.g., from an expression vector in a mammalian cell) and purified for pharmaceutical formulation.

The DNA expression constructs will typically include an expression control DNA sequence operably linked to the coding sequences, including naturally-associated or heterologous promoter regions. Preferably, the expression control sequences will be eukaryotic promoter systems in vectors capable of transforming or transfecting eukaryotic host cells. Once the vector has been incorporated into the appropriate host, the host is maintained under conditions suitable for high level expression of the

nucleotide sequences, and the collection and purification of the mutant' "engineered" antibodies.

As stated previously, the DNA sequences will be expressed in hosts after the sequences have been operably linked to an expression control sequence (i.e., positioned to ensure the transcription and translation of the structural gene). These expression vectors are typically replicable in the host organisms either as episomes or as an integral part of the host chromosomal DNA. Commonly, expression vectors will contain selection markers, e.g., tetracycline or neomycin, to permit detection of those cells transformed with the desired DNA sequences (see, e.g., **USPN 4,704,362**, which is incorporated herein by reference).

In addition to eukaryotic microorganisms such as yeast, mammalian tissue cell culture may also be used to produce the polypeptides of the present invention (see **Winnacker**, 1987), which is incorporated herein by reference). Eukaryotic cells are actually preferred, because a number of suitable host cell lines capable of secreting intact immunoglobulins have been developed in the art, and include the CHO cell lines, various COS cell lines, HeLa cells, and myeloma cell lines, but preferably transformed Bcells or hybridomas. Expression vectors for these cells can include expression control sequences, such as an origin of replication, a promoter, an enhancer (**Queen** et al, 1986), and necessary processing information sites, such as ribosome binding sites, RNA splice sites, polyadenylation sites, and transcriptional terminator sequences. Preferred expression control sequences are promoters derived from immunoglobulin genes, cytomegalovirus, SV40, Adenovirus, Bovine Papilloma Virus, and the like.

Eukaryotic DNA transcription can be increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting sequences of between 10 to 300 bp that increase transcription by a promoter. Enhancers can effectively increase transcription when either 51 or 31 to the transcription unit. They are also effective if

located within an intron or within the coding sequence itself. Typically, viral enhancers are used, including SV40 enhancers, cytomegalovirus enhancers, polyoma enhancers, and adenovirus enhancers. Enhancer sequences from mammalian systems are also commonly used, such as the mouse immunoglobulin heavy chain enhancer.

Mammalian expression vector systems will also typically include a selectable marker gene. Examples of suitable markers include, the dihydrofolate reductase gene (DHFR), the thymidine kinase gene (TK), or prokaryotic genes conferring drug resistance. The first two marker genes prefer the use of mutant cell lines that lack the ability to grow without the addition of thymidine to the growth medium. Transformed cells can then be identified by their ability to grow on non-supplemented media. Examples of prokaryotic drug resistance genes useful as markers include genes conferring resistance to G418, mycophenolic acid and hygromycin.

The vectors containing the DNA segments of interest can be transferred into the host cell by well-known methods, depending on the type of cellular host. For example, calcium chloride transfection is commonly utilized for prokaryotic cells, whereas calcium phosphate treatment. lipofection, or electroporation may be used for other cellular hosts. Other methods used to transform mammalian cells include the use of Polybrene, protoplast fusion, liposomes, electroporation, and micro-injection (see, generally, **Sambrook** et al, 1982 and 1989).

Once expressed, the antibodies, individual mutated immunoglobulin chains, mutated antibody fragments, and other immunoglobulin polypeptides of the invention can be purified according to standard procedures of the art, including ammonium sulfate precipitation, fraction column chromatography, gel electrophoresis and the like (see, generally, **Scopes**, 1982). Once purified, partially or to homogeneity as desired, the polypeptides may then be used therapeutically or in developing and performing assay

procedures, immunofluorescent stainings, and the like (see, generally, **Lefkovits** and **Pernis**, 1979 and 1981; **Lefkovits**, 1997).

The antibodies generated by the method of the present invention can be used for diagnosis and therapy. By way of illustration and not limitation, they can be used to treat cancer, autoimmune diseases, or viral infections. For treatment of cancer, the antibodies will typically bind to an antigen expressed preferentially on cancer cells, such as erbB-2, CEA, CD33, and many other antigens and binding members well known to those skilled in the art.

Two-Hybrid Based Screening Assays

Shuffling can also be used to recombinatorially diversify a pool of selected library members obtained by screening a two-hybrid screening system to identify library members which bind a predetermined polypeptide sequence. The selected library members are pooled and shuffled by *in vitro* and/or *in vivo* recombination. The shuffled pool can then be screened in a yeast two hybrid system to select library members which bind said predetermined polypeptide sequence (e. g., and SH2 domain) or which bind an alternate predetermined polypeptide sequence (e.g., an SH2 domain from another protein species).

An approach to identifying polypeptide sequences which bind to a predetermined polypeptide sequence has been to use a so-called "two-hybrid" system wherein the predetermined polypeptide sequence is present in a fusion protein (**Chien** et al, 1991). This approach identifies protein-protein interactions *in vivo* through reconstitution of a transcriptional activator (**Fields and Song**, 1989), the yeast Gal4 transcription protein. Typically, the method is based on the properties of the yeast Gal4 protein, which consists of separable domains responsible for DNA-binding and transcriptional activation.

Polynucleotides encoding two hybrid proteins, one consisting of the yeast Gal4 DNA-binding domain fused to a polypeptide sequence of a known protein and the other consisting of the Gal4 activation domain fused to a polypeptide sequence of a second protein, are constructed and introduced into a yeast host cell. Intermolecular binding between the two fusion proteins reconstitutes the Gal4 DNA-binding domain with the Gal4 activation domain, which leads to the transcriptional activation of a reporter gene (e.g., *lacz, HIS3)* which is operably linked to a Gal4 binding site. Typically, the two-hybrid method is used to identify novel polypeptide sequences which interact with a known protein (**Silver and Hunt**, 1993; **Durfee** et al, 1993; **Yang** et al, 1992; **Luban** et al, 1993; **Hardy** et al, 1992; **Bartel** et al, 1993; and **Vojtek** et al, 1993). However, variations of the two-hybrid method have been used to identify mutations of a known protein that affect its binding to a second known protein (**Li and Fields,** 1993; **Lalo** et al, 1993; **Jackson** et al, 1993; and **Madura** et al, 1993). Two-hybrid systems have also been used to identify interacting structural domains of two known proteins (**Bardwell** et al, 1993; **Chakrabarty** et al, 1992; **Staudinger** et al, 1993; and **Milne and Weaver** 1993) or domains responsible for oligomerization of a single protein (**Iwabuchi** et al, 1993; **Bogerd** et al, 1993). Variations of two-hybrid systems have been used to study the *in vivo* activity of a proteolytic enzyme (**Dasmahapatra** et al, 1992). Alternatively, an E. coli/BCCP interactive screening system (**Germino** et al, 1993; **Guarente,** 1993) can be used to identify interacting protein sequences (i.e., protein sequences which heterodimerize or form higher order heteromultimers). Sequences selected by a two-hybrid system can be pooled and shuffled and introduced into a two-hybrid system for one or more subsequent rounds of screening to identify polypeptide sequences which bind to the hybrid containing the predetermined binding sequence. The sequences thus identified can be compared to identify consensus sequence(s) and consensus sequence kernals.

In general, standard techniques of recombination DNA technology are described in various publications (e.g. **Sambrook** et al, 1989; **Ausubel** et al, 1987; and **Berger** and

Kimmel, 1987; each of which is incorporated herein in its entirety by reference. Polynucleotide modifying enzymes were used according to the manufacturer's recommendations. Oligonucleotides were synthesized on an Applied Biosystems Inc. Model 394 DNA synthesizer using ABI chemicals. If desired, PCR amplimers for amplifying a predetermined DNA sequence may be selected at the discretion of the practitioner.

One microgram samples of template DNA are obtained and treated with U.V. light to cause the formation of dimers, including TT dimers, particularly purine dimers. U.V. exposure is limited so that only a few photoproducts are generated per gene on the template DNA sample. Multiple samples are treated with U.V. light for varying periods of time to obtain template DNA samples with varying numbers of dimers from U.V. exposure.

A random priming kit which utilizes a non-proofreading polymease (for example, Prime-It II Random Primer Labeling kit by Stratagene Cloning Systems) is utilized to generate different size polynucleotides by priming at random sites on templates which are prepared by U.V. light (as described above) and extending along the templates. The priming protocols such as described in the Prime-It II Random Primer Labeling kit may be utilized to extend the primers. The dimers formed by U.V. exposure serve as a roadblock for the extension by the non-proofreading polymerase. Thus, a pool of random size polynucleotides is present after extension with the random primers is finished.

The present invention is further directed to a method for generating a selected mutant polynucleotide sequence (or a population of selected polynucleotide sequences) typically in the form of amplified and/or cloned polynucleotides, whereby the selected polynucleotide sequences(s) possess at least one desired phenotypic characteristic (e.g., encodes a polypeptide, promotes transcription of linked polynucleotides, binds a protein, and the like) which can be selected for. One method for identifying hybrid polypeptides

that possess a desired structure or functional property, such as binding to a predetermined biological macromolecule (e.g., a receptor), involves the screening of a large library of polypeptides for individual library members which possess the desired structure or functional property conferred by the amino acid sequence of the polypeptide.

In one embodiment, the present invention provides a method for generating libraries of displayed polypeptides or displayed antibodies suitable for affinity interaction screening or phenotypic screening. The method comprises (1) obtaining a first plurality of selected library members comprising a displayed polypeptide or displayed antibody and an associated polynucleotide encoding said displayed polypeptide or displayed antibody, and obtaining said associated polynucleotides or copies thereof wherein said associated polynucleotides comprise a region of substantially identical sequences, optimally introducing mutations into said polynucleotides or copies, (2) pooling the polynucleotides or copies, (3) producing smaller or shorter polynucleotides by interrupting a random or particularized priming and synthesis process or an amplification process, and (4) performing amplification, preferably PCR amplification, and optionally mutagenesis to homologously recombine the newly synthesized polynucleotides.

It is a particularly preferred object of the invention to provide a process for producing hybrid polynucleotides which express a useful hybrid polypeptide by a series of steps comprising:

(a) producing polynucleotides by interrupting a polynucleotide amplification or synthesis process with a means for blocking or interrupting the amplification or synthesis process and thus providing a plurality of smaller or shorter polynucleotides due to the replication of the polynucleotide being in various stages of completion;

(b) adding to the resultant population of single- or double-stranded polynucleotides one or more single- or double-stranded oligonucleotides, wherein said added oligonucleotides comprise an area of identity in an area of heterology to one or more of the single- or double-stranded polynucleotides of the population;

(c) denaturing the resulting single- or double-stranded oligonucleotides to produce a mixture of single-stranded polynucleotides, optionally separating the shorter or smaller polynucleotides into pools of polynucleotides having various lengths and further optionally subjecting said polynucleotides to a PCR procedure to amplify one or more oligonucleotides comprised by at least one of said polynucleotide pools;

(d) incubating a plurality of said polynucleotides or at least one pool of said polynucleotides with a polymerase under conditions which result in annealing of said single-stranded polynucleotides at regions of identity between the single-stranded polynucleotides and thus forming of a mutagenized double-stranded polynucleotide chain;

(e) optionally repeating steps (c) and (d);

(f) expressing at least one hybrid polypeptide from said polynucleotide chain, or chains; and

(g) screening said at least one hybrid polypeptide for a useful activity.

In a preferred aspect of the invention, the means for blocking or interrupting the amplification or synthesis process is by utilization of uv light, DNA adducts, DNA binding proteins.

In one embodiment of the invention, the DNA adducts, or polynucleotides comprising the DNA adducts, are removed from the polynucleotides or polynucleotide pool, such as by a process including heating the solution comprising the DNA fragments prior to further processing.

Having thus disclosed exemplary embodiments of the present invention, it should be noted by those skilled in the art that the disclosures are exemplary only and that various other alternatives, adaptations and modifications may be made within the scope of the present invention. Accordingly, the present invention is not limited to the specific embodiments as illustrated herein.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following examples are to be considered illustrative and thus are not limiting of the remainder of the disclosure in any way whatsoever.

Example 1

Generation of Random Size Polynucleotides Using U.V. Induced Photoproducts

One microgram samples of template DNA are obtained and treated with U.V. light to cause the formation of dimers, including TT dimers, particularly purine dimers. U.V. exposure is limited so that only a few photoproducts are generated per gene on the template DNA sample. Multiple samples are treated with U.V. light for varying periods of time to obtain template DNA samples with varying numbers of dimers from U.V. exposure.

A random priming kit which utilizes a non-proofreading polymerase (for example, Prime-It II Random Primer Labeling kit by Stratagene Cloning Systems) is utilized to generate different size polynucleotides by priming at random sites on templates which are prepared by U.V. light (as described above) and extending along the templates. The priming protocols such as described in the Prime-It II Random Primer Labeling kit may be utilized to extend the primers. The dimers formed by U.V. exposure serve as a roadblock for the extension by the non-proofreading polymerase. Thus, a pool of random size polynucleotides is present after extension with the random primers is finished.

Example 2

Isolation of Random Size Polynucleotides

Polynucleotides of interest which are generated according to Example 1 are gel isolated on a 1.5% agarose gel. Polynucleotides in the 100-300 bp range are cut out of the gel and 3 volumes of 6 M NaI is added to the gel slice. The mixture is incubated at 50 °C for 10 minutes and 10 µl of glass milk (Bio 101) is added. The mixture is spun for 1 minute and the supernatant is decanted. The pellet is washed with 500 µl of Column Wash (Column Wash is 50% ethanol, 10mM Tris-HCl pH 7.5, 100 mM NaCl and 2.5 mM EDTA) and spin for 1 minute, after which the supernatant is decanted. The washing, spinning and decanting steps are then repeated. The glass milk pellet is resuspended in 20µl of H_2O and spun for 1 minute. DNA remains in the aqueous phase.

Example 3

Shuffling of Isolated Random Size 100-300bp Polynucleotides

The 100-300 bp polynucleotides obtained in Example 2 are recombined in an annealing mixture (0.2 mM each dNTP, 2.2 mM $MgCl_2$, 50 mM KCl, 10 mM Tris-HCl ph 8.8, 0.1% Triton X-100, 0.3 μ; Taq DNA polymerase, 50 μl total volume) without adding primers. A Robocycler by Stratagene was used for the annealing step with the following program: 95 °C for 30 seconds, 25-50 cycles of [95 °C for 30 seconds, 50 - 60 °C (preferably 58 °C) for 30 seconds, and 72 °C for 30 seconds] and 5 minutes at 72 °C. Thus, the 100-300 bp polynucleotides combine to yield double-stranded polynucleotides having a longer sequence. After separating out the reassembled double-stranded polynucleotides and denaturing them to form single stranded polynucleotides, the cycling is optionally again repeated with some samples utilizing the single strands as template and primer DNA and other samples utilizing random primers in addition to the single strands.

Example 4

Screening of Polypeptides from Shuffled Polynucleotides

The polynucleotides of Example 3 are separated and polypeptides are expressed therefrom. The original template DNA is utilized as a comparative control by obtaining comparative polypeptides therefrom. The polypeptides obtained from the shuffled polynucleotides of Example 3 are screened for the activity of the polypeptides obtained from the original template and compared with the activity levels of the control. The shuffled polynucleotides coding for interesting polypeptides discovered during screening are compared further for secondary desirable traits. Some shuffled polynucleotides corresponding to less interesting screened polypeptides are subjected to reshuffling.

Example 5

Directed Evolution an Enzyme by Saturation Mutagenesis

Site-Saturation Mutagenesis: To accomplish site-saturation mutagenesis every residue (316) of a dehalogenase enzyme was converted into all 20 amino acids by site directed mutagenesis using 32-fold degenerate oligonucleotide primers, as follows:

1. A culture of the dehalogenase expression construct was grown and a preparation of the plasmid was made
2. Primers were made to randomize each codon - they have the common structure $X_{20}NN(G/T)X_{20}$
3. A reaction mix of 25 ul was prepared containing ~50 ng of plasmid template, 125 ng of each primer, 1X native Pfu buffer, 200 uM each dNTP and 2.5 U native Pfu DNA polymerase
4. The reaction was cycled in a Robo96 Gradient Cycler as follows:
   Initial denaturation at 95°C for 1 min
   20 cycles of 95°C for 45 sec, 53°C for 1 min and 72°C for 11 min
   Final elongation step of 72°C for 10 min
5. The reaction mix was digested with 10 U of DpnI at 37°C for 1 hour to digest the methylated template DNA
6. Two ul of the reaction mix were used to transform 50 ul of XL1-Blue MRF' cells and the entire transformation mix was plated on a large LB-Amp-Met plate yielding 200-1000 colonies
7. Individual colonies were toothpicked into the wells of 96-well microtiter plates containing LB-Amp-IPTG and grown overnight
8. The clones on these plates were assayed the following day

Screening: Approximately 200 clones of mutants for each position were grown in liquid media (384 well microtiter plates) and screened as follows:

1. Overnight cultures in 384-well plates were centrifuged and the media removed. To each well was added 0.06 mL 1 mM Tris/SO_4^{2-} pH 7.8.
2. Made 2 assay plates from each parent growth plate consisting of 0.02 mL cell suspension.
3. One assay plate was placed at room temperature and the other at elevated temperature (initial screen used 55°C) for a period of time (initially 30 minutes).
4. After the prescribed time 0.08 mL room temperature substrate (TCP saturated 1 mM Tris/SO_4^{2-} pH 7.8 with 1.5 mM NaN_3 and 0.1 mM bromothymol blue) was added to each well.
5. Measurements at 620 nm were taken at various time points to generate a progress curve for each well.
6. Data were analyzed and the kinetics of the cells heated to those not heated were compared. Each plate contained 1-2 columns (24 wells) of unmutated 20F12 controls.
7. Wells that appeared to have improved stability were re-grown and tested under the same conditions.

Following this procedure nine single site mutations appeared to confer increased thermal stability on the enzyme. Sequence analysis was performed to determine of the exact amino acid changes at each position that were specifically responsible for the improvement. In sum, the improvement was conferred at 7 sites by one amino acid change alone, at an eighth site by each of two amino acid changes, and at a ninth site by each of three amino acid changes. Several mutants were then made each having a

plurality of these nine beneficial site mutations in combination; of these two mutants proved superior to all the other mutants, including those with single point mutations.

Example 6

Direct expression cloning using end-selection

An esterase gene was amplified using 5'phosphorylated primers in a standard PCR reaction (10 ng template; PCR conditions: 3' 94 C; [1' 94 C; 1' 50 C; 1'30" 68 C] x 30; 10' 68 C.

Forward Primer = 9511TopF
(CTAGAAGGGAGGAGAATTACATGAAGCGGCTTTTAGCCC)
Reverse Primer = 9511TopR (AGCTAAGGGTCAAGGCCGCACCCGAGG)
The resulting PCR product (ca.1000 bp) was gel purified and quantified.

A vector for expression cloning, pASK3 (Institut fuer Bioanalytik, Goettingen, Germany), was cut with *Xba* I and *Bgl* II and dephosphorylated with CIP.

0.5 pmoles Vaccina Topoisomerase I (Invitrogen, Carlsbad, CA) was added to 60 ng (ca. 0.1 pmole) purified PCR product for 5' 37 C in buffer NEB I (New England Biolabs, Beverly, MA) in 5 µl total volume.
The topogated PCR product was cloned into the vector pASK3 (5 µl, ca. 200 ng in NEB I) for 5' at room temperature.
This mixture was dialyzed against H_2O for 30'.
2 µl were used for electroporation of DH10B cells (Gibco BRL, Gaithersburg, MD).

Efficiency: Based on the actual clone numbers this method can produce 2 x 10^6 clones per µg vector. All tested recombinants showed esterase activity after induction with anhydrotetracycline.

Example 7

Dehalogenase Thermal Stability

This invention provides that a desirable property to be generated by directed evolution is exemplified in a limiting fashion by an improved residual activity (e.g. an enzymatic activity, an immunoreactivity, an antibiotic acivity, etc.) of a molecule upon subjection to altered environment, including what may be considered a harsh enviroment, for a specified time. Such a harsh environment may comprise any combination of the following (iteratively or not, and in any order or permutation): an elevated temperature (including a temperature that may cause denaturation of a working enzyme), a decreased temperature, an elevated salinity, a decreased salinity, an elevated pH, a decreased pH, an elevated pressure, a decreassed pressure, and an change in exposure to a radiation source (including uv radiation, visible light, as well as the entire electromagnetic spectrum).

The following example shows an application of directed evolution to evolve the ability of an enzyme to regain &/or retain activity upon exposure to an elevated temperature.

Every residue (316) of a dehalogenase enzyme was converted into all 20 amino acids by site directed mutagenesis using 32-fold degenerate oligonucleotide primers. These mutations were introduced into the already rate-improved variant Dhla 20F12. Approximately 200 clones of each position were grown in liquid media (384 well microtiter plates) to be screened. The screening procedure was as follows:

1. Overnight cultures in 384-well plates were centrifuged and the media removed. To each well was added 0.06 mL 1 mM Tris/SO_4^{2-} pH 7.8.
2. The robot made 2 assay plates from each parent growth plate consisting of 0.02 mL cell suspension.
3. One assay plate was placed at room temperature and the other at elevated temperature (initial screen used 55°C) for a period of time (initially 30 minutes).

4. After the prescribed time 0.08 mL room temperature substrate (TCP saturated 1 mM Tris/SO_4^{2-} pH 7.8 with 1.5 mM NaN_3 and 0.1 mM bromothymol blue) was added to each well. TCP = trichloropropane.
5. Measurements at 620 nm were taken at various time points to generate a progress curve for each well.
6. Data were analyzed and the kinetics of the cells heated to those not heated were compared. Each plate contained 1-2 columns (24 wells) of un-mutated 20F12 controls.
7. Wells that appeared to have improved stability were regrown and tested under the same conditions.

Following this procedure nine single site mutations appeared to confer increased thermal stability on Dhla-20F12. Sequence analysis showed that the following changes were beneficial:

D89G
F91S
T159L
G189Q, G189V
I220L
N238T
W251Y
P302A, P302L, P302S, P302K
P302R/S306R

Only two sites (189 and 302) had more than one substitution. The first 5 on the list were combined (using G189Q) into a single gene (this mutant is referred to as "Dhla5"). All changes but S306R were incorporated into another variant referred to as Dhla8.

Thermal stability was assessed by incubating the enzyme at the elevated temperature (55°C and 80°C) for some period of time and activity assay at 30°C. Initial rates were plotted vs. time at the higher temperature. The enzyme was in 50 mM Tris/SO_4 pH 7.8 for both the incubation and the assay. Product (Cl^-) was detected by a standard method using $Fe(NO_3)_3$ and HgSCN. Dhla 20F12 was used as the *de facto* wild type. The apparent half-life ($T_{1/2}$) was calculated by fitting the data to an exponential decay function.

3. LITERATURE CITED

Unless otherwise indicated, all references cited herein (supra and infra) are incorporated by reference in their entirety.

Barret AJ, et al., eds.: Enzyme Nomenclature: Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology. San Diego: Academic Press, Inc., 1992.

Boyce COL, ed.: Novo's Handbook of Practical Biotechnology. 2nd ed. Bagsvaerd, Denmark, 1986.

Drauz K, Waldman H, eds.: Enzyme Catalysis in Organic Synthesis: A Comprehensive Handbook. Vol. 1. New York: VCH Publishers, 1995.

Drauz K, Waldman H, eds.: Enzyme Catalysis in Organic Synthesis: A Comprehensive Handbook. Vol. 2. New York: VCH Publishers, 1995.

Foster GD, Taylor SC, eds.: Plant Virology Protocols: From Virus Isolation to Transgenic Resistance. Methods in Molecular Biology, Vol. 81. New Jersey: Humana Press Inc., 1998.

Franks F, ed.: Protein Biotechnology: Isolation, Characterization, and Stabilization. New Jersey: Humana Press Inc., 1993.

Godfrey T, West S, eds.: Industrial Enzymology. 2nd ed. London: Macmillan Press Ltd, 1996.

Gottschalk G: Bacterial Metabolism. 2nd ed. New York: Springer-Verlag Inc., 1986.

Gresshoff PM, ed.: Technology Transfer of Plant Biotechnology. Current Topics in Plant Molecular Biology. Boca Raton: CRC Press, 1997.

Griffin HG, Griffin AM, eds.: PCR Technology: Current Innovations. Boca Raton: CRC Press, Inc., 1994.

Hansen G, Chilton MD: Lessons in gene transfer to plants by a gifted microbe. *Curr Top Microbiol Immunol* 240:21-57, 1999.

Hartmann HT, et al.: Plant Propagation: Principles and Practices. 6th ed. New Jersey: Prentice Hall, Inc., 1997.

Perun TJ, Propst CL, eds.: Computer-Aided Drug Design: Methods and Applications. New York: Marcel Dekker, Inc., 1989.

Owen MRL, Pen J: Transgenic Plants: A Production System for Industrial and Pharmaceutical Proteins. Chichester: John Wiley & Sons, 1996.

Segel IH: Enzyme Kinetics: Behavior and Analysis of Rapid Equilibrium and Steady-State Enzyme Systems. New York: John Wiley & Sons, Inc., 1993.

White JS, White DC: Source Book of Enzymes. Boca Raton: CRC Press, 1997.

Wong CH, Whitesides GM: Enzymes in Synthetic Organic Chemistry. Vol. 12. New York: Elsevier Science Publications, 1995.

WO 97/35966; Filed Mar. 20, 1997, Published Oct. 2, 1997. Minshull J, Stemmer WP: Mehtods and Compositions for Cellular and Metabolic Engineering.

WO 98/31837; Filed Jan. 16, 1998, Published July 23, 1998. Delcardayre SB, Tobin MB, Stemmer WP, Minshull, J: Evolution of Whole Cells and Organisms by Recursive Sequence Recombination.

WO 98/37223; Filed Feb. 18 1998, Published Aug. 27, 1998. Pang SZ, Gonsalves D, Jan FJ: DNA Construct to Confer Multiple Traits on Plants.

Alting-Mecs MA and Short JM: Polycos vectors: a system for packaging filamentous phage and phagemid vectors using lambda phage packaging extracts. *Gene* 137:1, 93-100, 1993.

Arkin AP and Youvan DC: An algorithm for protein engineering: simulations of recursive ensemble mutagenesis. *Proc Natl Acad Sci USA* 89(16):7811-7815, (Aug 15) 1992.

Arnold FH: Protein engineering for unusual environments. *Current Opinion in Biotechnology* 4(4):450-455, 1993.

Ausubel FM, et al Editors. Current Protocols in Molecular Biology, Vols. 1 and 2 and supplements. (a.k.a. "The Red Book") Greene Publishing Assoc., Brooklyn, NY, ©1987.

Ausubel FM, et al Editors. Current Protocols in Molecular Biology, Vols. 1 and 2 and supplements. (a.k.a. "The Red Book") Greene Publishing Assoc., Brooklyn, NY, ©1989.

Ausubel FM, et al Editors. Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology. Greene Publishing Assoc., Brooklyn, NY, ©1989.

Ausubel FM, et al Editors. Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology, 2nd Edition. Greene Publishing Assoc., Brooklyn, NY, ©1992.

Barbas CF 3d, Bain JD, Hoekstra DM, Lerner RA: Semisynthetic combinatorial antibody libraries: a chemical solution to the diversity problem. *Proc Natl Acad Sci USA* 89(10):4457-4461, 1992.

Bardwell AJ, Bardwell L, Johnson DK, Friedberg EC: Yeast DNA recombination and repair proteins Rad1 and Rad10 constitute a complex in vivo mediated by localized hydrophobic domains. *Mol Microbiol* 8(6):1177-1188, 1993.

Bartel P, Chien CT, Sternglanz R, Fields S: Elimination of false positives that arise in using the two-hybrid system. *Biotechniques* 14(6):920-924, 1993.

Beaudry AA and Joyce GF: Directed evolution of an RNA enzyme. *Science* 257(5070):635-641, 1992.

Berger and Kimmel, Methods in Enzymology, Volume 152, Guide to Molecular Cloning Techniques. Academic Press, Inc., San Diego, CA, ©1987. (Cumulative Subject Index: Volumes 135-139, 141-167, 1990, 272 pp.)

Bevan M: Binary Agrobacterium vectors for plant transformation. *Nucleic Acids Research* 12(22):8711-21, 1984.

Biocca S, Pierandrei-Amaldi P, Cattaneo A: Intracellular expression of anti-p21ras single chain Fv fragments inhibits meiotic maturation of xenopus oocytes. *Biochem Biophys Res Commun* 197(2):422-427, 1993.

Bird et al. *Plant Mol Biol* 11:651, 1988..

Bogerd HP, Fridell RA, Blair WS, Cullen BR: Genetic evidence that the Tat proteins of human immunodeficiency virus types 1 and 2 can multimerize in the eukaryotic cell nucleus. *J Virol* 67(8):5030-5034, 1993.

Brederode FT, Koper-Zawrthoff EC, Bol JF: Complete nucleotide sequence of alfalfa mosaic virus RNA 4. *Nucleic Acids Research* 8(10):2213-23, 1980.

Breitling F, Dubel S, Seehaus T, Klewinghaus I, Little M: A surface expression vector for antibody screening. *Gene* 104(2):147-153, 1991.

Brown NL, Smith M: Cleavage specificity of the restriction endonuclease isolated from Haemophilus gallinarum (Hga I). *Proc Natl Acad Sci U S A* 74(8):3213-6, (Aug) 1977.

Burton DR, Barbas CF 3d, Persson MA, Koenig S, Chanock RM, Lerner RA: A large array of human monoclonal antibodies to type 1 human immunodeficiency virus from combinatorial libraries of asymptomatic seropositive individuals. *Proc Natl Acad Sci U S A* 88(22):10134-7, (Nov 15) 1991.

Caldwell RC and Joyce GF: Randomization of genes by PCR mutagenesis. *PCR Methods Appl* 2(10):28-33, 1992.

Caton AJ and Koprowski H: Influenze virus hemagglutinin-specific antibodies isolatedf froma combinatorial expression library are closely related to the immune response of the donor. *Proc Natl Acad Sci USA* 87(16):6450-6454, 1990.

Chakraborty T, Martin JF, Olson EN: Analysis of the oligomerization of myogenin and E2A products in vivo using a two-hybrid assay system. *J Biol Chem* 267(25):17498-501, 1992.

Chang CN, Landolfi NF, Queen C: Expression of antibody Fab domains on bacteriophage surfaces. Potential use for antibody selection. *J Immunol* 147(10):3610-4, (Nov 15) 1991.

Chaudhary VK, Batra JK, Gallo MG, Willingham MC, FitzGerald DJ, Pastan I: A rapid method of cloning functional variable-region antibody genes in Escherichia coli as single-chain immunotoxins. *Proc Natl Acad Sci USA* 87(3):1066-1070, 1990.

Chien CT, Bartel PL, Sternglanz R, Fields S: The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest. *Proc Natl Acad Sci USA* 88(21):9578-9582, 1991.

Chiswell DJ, McCafferty J: Phage antibodies: will new 'coliclonal' antibodies replace monoclonal antibodies? *Trends Biotechnol* 10(3):80-84, 1992.

Chothia C and Lesk AM: Canonical structures for the hypervariable regions of immunoglobulins. *J Mol Biol* 196)4):901-917, 1987.

Chothia C, Lesk AM, Tramontano A, Levitt M, Smith-Gill SJ, Air G, Sheriff S, Padlan EA, Davies D, Tulip WR, et al: Conformations of immunoglobulin hypervariable regions. *Nature* 342(6252):877-883, 1989.

Clackson T, Hoogenboom HR, Griffiths AD, Winter G: Making antibody fragments using phage display libraries. *Nature* 352(6336):624-628, 1991.

Conrad M, Topal MD: DNA and spermidine provide a switch mechanism to regulate the activity of restriction enzyme *Nae* I. *Proc Natl Acad Sci U S A* 86(24):9707-11, (Dec) 1989.

Coruzzi G, Broglie R, Edwards C, Chua NH: Tissue-specific and light-regulated expression of a pea nuclear gene encoding the small subunit of ribulose-1,5-bisphosphate carboxylase. *EMBO J* 3(8):1671-9, 1984.

Dasmahapatra B, DiDomenico B, Dwyer S, Ma J, Sadowski I, Schwartz J: A genetic system for studying the activity of a proteolytic enzyme. *Proc Natl Acad Sci USA* 89(9):4159-4162, 1992.

Davis LG, Dibner MD, Battey JF. Basic Methods in Molecular Biology. Elsevier, New York, NY, ©1986.

Delegrave S and Youvan DC. *Biotechnology Research* 11:1548-1552, 1993.

DeLong EF, Wu KY, Prezelin BB, Jovine RV: High abundance of Archaea in Antarctic marine picoplankton. *Nature* 371(6499):695-697, 1994.

Deng SJ, MacKenzie CR, Sadowska J, Michniewicz J, Young NM, Bundle Dr, Narang SA: Selection of antibody single-chain variable fragments with improved carbohydrate binding by phage display. *J Biol Chem* 269(13):9533-9538, 1994.

Duan L, Bagasra O, Laughlin MA, Oakes JW, Pomerantz RJ: Potent inhibition of human immunodeficiency virus type 1 replication by an intracellular anti-Rev single-chain antibody. *Proc Natl Acad Sci USA* 91(11):5075-5079, 1994.

Durfee T, Becherer K, Chen PL, Yeh SH, Yang Y, Kilburn AE, Lee WH, Elledge SJ: The retinoblastoma protein associates with the protein phosphatase type 1 catalytic subunit. *Genes Dev* 7(4):555-569, 1993.

Ellington AD and Szostak JW: In vitro selection of RNA molecules that bind specific ligands. *Nature* 346(6287):818-822, 1990.

Fields S and Song 0: A novel genetic system to detect protein-protein interactions. *Nature* 340(6230):245-246, 1989.

Firek S, Draper J, Owen MR, Gandecha A, Cockburn B, Whitelam GC: Secretion of a functional single-chain Fv protein in transgenic tobacco plants and cell suspension cultures. *Plant Mol Biol* 23(4):861-870, 1993.

Forsblom S, Rigler R, Ehrenberg M, Philipson L: Kinetic studies on the cleavage of adenovirus DNA by restriction endonuclease Eco RI. *Nucleic Acids Res* 3(12):3255-69, (Dec) 1976.

Germino FJ, Wang ZX, Weissman SM: Screening for in vivo protein-protein interactions. *Proc Natl Acad Sci USA* 90(3):933-937, 1993.

Gingeras TR, Brooks JE: Cloned restriction/modification system from Pseudomonas aeruginosa. *Proc Natl Acad Sci USA* 80(2):402-6, 1983 (Jan).

Gluzman Y: SV40-transformed simian cells support the replication of early SV40 mutants. *Cell* 23(1):175-182, 1981.

Gruber M, Schodin BA, Wilson ER, Kranz DM: Efficient tumor cell lysis mediated by a bispecific single chain antibody expressed in Escherichia coli. *J Immunol* 152(11):5368-5374, 1994.

Guarente L: Strategies for the identification of interacting proteins. *Proc Natl Acad Sci USA* 90(5):1639-1641, 1993.

Guilley H, Dudley RK, Jonard G, Balazs E, Richards KE: Transcription of Cauliflower mosaic virus DNA: detection of promoter sequences, and characterization of transcripts. *Cell* 30(3):763-73, 1982.

Hardy CF, Sussel L, Shore D: A RAP1-interacting protein involved in transcriptional silencing and telomere length regulation. *Genes Dev* 6(5):801-814, 1992.

Hawkins RE and Winter G: Cell selection strategies for making antibodies from variable gene libraries: trapping the memory pool. *Eur J Immunol* 22(3):867-870, 1992.

Holvoet P, Laroche Y, Lijnen HR, Van Hoef B, Brouwers E, De Cock F, Lauwereys M, Gansemans Y, Collen D: Biochemical characterization of single-chain chimeric plasminogen activators consisting of a single-chain Fv fragment of a fibrin-specific antibody and single-chain urokinase. *Eur J Biochem* 210(3):945-952, 1992.

Honjo T, Alt FW, Rabbitts TH (eds): *Immunoglobulin genes*. Academic Press: San Diego, CA, pp. 361-368, ©1989.

Hoogenboom HR, Griffiths AD, Johnson KS, Chiswell DJ, Judson P, Winter G: Multi-subunit proteins on the surface of filamentous phage: methodologies for displaying antibody (Fab) heavy and light chains. *Nucleic Acids Res* 19(15):4133-4137, 1991.

Huse WD, Sastry L, Iverson SA, Kang AS, Alting-Mees M, Burton DR, Benkovic SJ, Lerner RA: Generation of a large combinatorial library of the immunoglobulin repertoire in phage lambda. *Science* 246(4935):1275-1281, 1989.

Huston JS, Levinson D, Mudgett-Hunter M, Tai MS, Novotney J, Margolies MN, Ridge RJ, Bruccoleri RE, Haber E, Crea R, et al: Protein engineering of antibody binding sites: recovery of specific activity in an anti-digoxin single-chain Fv analogue produced in Escherichia coli. *Proc Natl Acad Sci USA* 85(16):5879-5883, 1988.

Iwabuchi K, Li B, Bartel P, Fields S: Use of the two-hybrid system to identify the domain of p53 involved in oligomerization. *Oncogene* 8(6):1693-1696, 1993.

Jackson AL, Pahl PM, Harrison K, Rosamond J, Sclafani RA: Cell cycle regulation of the yeast Cdc7 protein kinase by association with the Dbf4 protein. *Mol Cell Biol* 13(5):2899-2908, 1993.

Johnson S and Bird RE: *Methods Enzymol* 203:88, 1991.

Kabat et al: Sequences of Proteins of Immunological Interest, 4th Ed. U.S. Department of Health and Human Services, Bethesda, MD (1987)

Kang AS, Barbas CF, Janda KD, Benkovic SJ, Lerner RA: Linkage of recognition and replication functions by assembling combinatorial antibody Fab libraries along phage surfaces. *Proc Natl Acad Sci USA* 88(10):4363-4366, 1991.

Kettleborough CA, Ansell KH, Allen RW, Rosell-Vives E, Gussow DH, Bendig MM: Isolation of tumor cell-specific single-chain Fv from immunized mice using phage-antibody libraries and the re-construction of whole antibodies from these antibody fragments. *Eur J Immunol* 24(4):952-958, 1994.

Kruger DH, Barcak GJ, Reuter M, Smith HO: EcoRII can be activated to cleave refractory DNA recognition sites. *Nucleic Acids Res* 16(9):3997-4008, (May 11) 1988.

Lalo D, Carles C, Sentenac A, Thuriaux P: Interactions between three common subunits of yeast RNA polymerases I and III. *Proc Natl Acad Sci USA* 90(12):5524-5528, 1993.

Laskowski M Sr: Purification and properties of venom phosphodiesterase. *Methods Enzymol* 65(1):276-84, 1980.

Lefkovits I and Pernis B, Editors. Immunological Methods, Vols. I and II. Academic Press, New York, NY. Also Vol. III published in Orlando and Vol. IV published in San Diego. ©1979-.

Ivan Lefkovits, Editor. Immunology methods manual : the comprehensive sourcebook of techniques. Academic Press, San Diego, ©1997.

Lerner RA, Kang AS, Bain JD, Burton DR, Barbas CF 3d: Antibodies without immunization. *Science* 258(5086):1313-1314, 1992.

Leung, D.W., et al, *Technique*, 1:11-15, 1989.

Li B and Fields S: Identification of mutations in p53 that affect its binding to SV40 large T antigen by using the yeast two-hybrid system. *FASEB J* 7(10):957-963, 1993.

Lilley GG, Doelzal O, Hillyard CJ, Bernard C, Hudson PJ: Recombinant single-chain antibody peptide conjugates expressed in Escherichia coli for the rapid diagnosis of HIV. *J Immunol Methods* 171(2):211-226, 1994.

Lowman HB, Bass SH, Simpson N, Wells JA: Selecting high-affinity binding proteins by monovalent phage display. *Biochemistry* 30(45):10832-10838, 1991.

Luban J, Bossolt KL, Franke EK, Kalpana GV, Goff SP: Human immunodeficiency virus type 1 Gag protein binds to cyclophilins A and B. *Cell* 73(6):1067-1078, 1993.

Madura K, Dohmen RJ, Varshavsky A: N-recognin/Ubc2 interactions in the N-end rule pathway. *J Biol Chem* 268(16):12046-54, (Jun 5) 1993.

Marks JD, Hoogenboom HR, Bonnert TP, McCafferty J, Griffiths AD, Winter G: By-passing immunization. Human antibodies from V-gene libraries displayed on phage. *J Mol Biol* 222(3):581-597, 1991.

Marks JD, Griffiths Ad, Malmqvist M, Clackson TP, Bye JM, Winter G: By-passing immunization: building high affinity human antibodies by chain shuffling. *Biotechnology (N Y)* 10(7):779-783, 1992.

Marks JD, Hoogenboom HR, Griffiths AD, Winter G: Molecular evolution of proteins on filamentous phage. Mimicking the strategy of the immune system. *J Biol Chem* 267(23):16007-16010, 1992.

Maxam AM, Gilbert W: Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol* 65(1):499-560, 1980.

McCafferty J, Griffiths AD, Winter G, Chiswell DJ: Phage antibodies: filamentous phage displaying antibody variable domains. *Nature* 348(6301):552-554, 1990.

Miller JH. A Short Course in Bacterial Genetics: A Laboratory Manual and Handbook for Escherichia coli and Related Bacteria (see inclusively p. 445). Cold Spring Harbor Laboratory Press, Plainview, NY, ©1992.

Milne GT and Weaver DT: Dominant negative alleles of RAD52 reveal a DNA repair/ recombination complex including Rad51 and Rad52. *Genes Dev* 7(9):1755-1765, 1993.

Mullinax RL, Gross EA, Amberg JR, Hay BN, Hogrefe HH, Kubtiz MM, Greener A, Alting-Mees M, Ardourel D, Short JM, et al: Identification of human antibody fragment clones specific for tetanus toxoid in a bacteriophage lambda immunoexpression library. *Proc natl Acad Sci USA* 87(20):8095-9099, 1990.

Nath K, Azzolina BA: in *Gene Amplification and Analysis* (ed. Chirikjian JG), vol. 1, p. 113, Elsevier North Holland, Inc., New York, New York, ©1981.

Needleman SB and Wunsch CD: A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J Mol Biol* 48(3):443-453, 1970.

Nelson M, Christ C, Schildkraut I: Alteration of apparent restriction endonuclease recognition specificities by DNA methylases. *Nucleic Acids Res* 12(13):5165-73, 1984 (Jul 11).

Nicholls PJ, Johnson VG, Andrew SM, Hoogenboom HR, Raus JC, Youle RJ: Characterization of single-chain antibody (sFv)-toxin fusion proteins produced in vitro in rabbit reticulocyte lysate. *J Biol Chem* 268(7):5302-5308, 1993.

Oller AR, Vanden Broek W, Conrad M, Topal MD: Ability of DNA and spermidine to affect the activity of restriction endonucleases from several bacterial species. *Biochemistry* 30(9):2543-9, (Mar 5) 1991.

Owens RJ and Young RJ: The genetic engineering of monoclonal antibodies. *J Immunol Methods* 168(2):149-165, 1994.

Pearson WR and Lipman DJ: Improved tools for biological sequence comparison. *Proc Natl Acad Sci USA* 85(8):2444-2448, 1988.

Pein CD, Reuter M, Meisel A, Cech D, Kruger DH: Activation of restriction endonuclease EcoRII does not depend on the cleavage of stimulator DNA. *Nucleic Acids Re s* 19(19):5139-42, (Oct 11) 1991.

Persson MA, Caothien RH, Burton DR: Generation of diverse high-affinity human monoclonal antibodies by repertoire cloning. *Proc Natl Acad Sci USA* 88(6):2432-2436, 1991.

Queen C, Foster J, Stauber C, Stafford J: Cell-type specific regulation of a kappa immunoglobulin gene by promoter and enhance elements. *Immunol Rev* 89:49-68, 1986.

Qiang BQ, McClelland M, Poddar S, Spokauskas A, Nelson M: The apparent specificity of NotI (5'-GCGGCCGC-3') is enhanced by M.FnuDII or M.BepI methyltransferases (5'-mCGCG-3'): cutting bacterial chromosomes into a few large pieces. *Gene* 88(1):101-5, (Mar 30) 1990.

Raleigh EA, Wilson G: Escherichia coli K-12 restricts DNA containing 5-methylcytosine. *Proc Natl Acad Sci U S A* 83(23):9070-4, (Dec) 1986.

Reidhaar-Olson JF and Sauer RT: Combinatorial cassette mutagenesis as a probe of the informational content of protein sequences. *Science* 241(4861):53-57, 1988.

Riechmann L and Weill M: Phage display and selection of a site-directed randomized single-chain antibody Fv fragment for its affinity improvement. *Biochemistry* 32(34):8848-8855, 1993.

Roberts RJ, Macelis D: REBASE--restriction enzymes and methylases. *Nucleic Acids Res* 24(1):223-35, (Jan 1) 1996.

Ryan AJ, Royal CL, Hutchinson J, Shaw CH: Genomic sequence of a 12S seed storage protein from oilseed rape (Brassica napus c.v. jet neuf). *Nucl Acids Res* 17(9):3584, 1989.

Sambrook J, Fritsch EF, Maniatis T. Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, ©1982.

Sambrook J, Fritsch EF, Maniatis T. Molecular Cloning: A Laboratory Manual. Second Edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, ©1989.

Scopes RK. Protein Purification: Principles and Practice. Springer-Verlag, New York, NY, © 1982.

Silver SC and Hunt SW 3d: Techniques for cloning cDNAs encoding interactive transcriptional regulatory proteins. *Mol Biol Rep* 17(3):155-165, 1993.

Smith TF, Waterman MS. *Adv Appl Math* 2: 482-end of article, 1981.

Smith TF, Waterman MS: Overlapping genes and information theory. *J Theor Biol* 91(2):379-80, (Jul 21) 1981.

Smith TF, Waterman MS: Identification of common molecular subsequences. *J Mol Biol* 147(1):195-7, (Mar 25) 1981.

Smith TF, Waterman MS, Fitch WM: Comparative biosequence metrics. *J Mol Evol* S18(1):38-46, 1981.

Staudinger J, Perry M, Elledge SJ, Olson EN: Interactions among vertebrate helix-loop-helix proteins in yeast using the two-hybrid system. *J Biol Chem* 268(7):4608-4611, 1993.

Stemmer WP, Morris SK, Wilson BS: Selection of an active single chain Fv antibody from a protein linker library prepared by enzymatic inverse PCR. *Biotechniques* 14(2):256-265, 1993.

Stemmer WP: DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proc Natl Acad Sci USA* 91(22):10747-10751, 1994.

Sun D, Hurley LH: Effect of the (+)-CC-1065-(N3-adenine)DNA adduct on in vitro DNA synthesis mediated by Escherichia coli DNA polymerase. *Biochemistry* 31:10, 2822-9, (Mar 17) 1992,

Tague BW, Dickinson CD, Chrispeels MJ: A short domain of the plant vacuolar protein phytohemagglutinin targets invertase to the yeast vacuole. *Plant Cell* 2(6):533-46, (June) 1990.

Takahashi N, Kobayashi I: Evidence for the double-strand break repair model of bacteriophage lambda recombination. *Proc Natl Acad Sci U S A* 87(7):2790-4, (Apr) 1990.

Thiesen HJ and Bach C: Target Detection Assay (TDA): a versatile procedure to determine DNA binding sites as demonstrated on SP1 protein. *Nucleic Acids Res* 18(11):3203-3209, 1990.

Thomas M, Davis RW: Studies on the cleavage of bacteriophage lambda DNA with EcoRI Restriction endonuclease. *J Mol Biol* 91(3):315-28, (Jan 25) 1975.

Tingey SV, Walker EL, Corruzzi GM: Glutamine synthetase genes of pea encode distinct polypeptides which are differentially expressed in leaves, roots and nodules. *EMBO J* 6(1):1-9, 1987.

Topal MD, Thresher RJ, Conrad M, Griffith J: NaeI endonuclease binding to pBR322 DNA induces looping. *Biochemistry* 30(7):2006-10, (Feb. 19) 1991.

Tramontano A, Chothia C, Lesk AM: Framework residue 71 is a major determinant of the position and conformation of the second hypervariable region in the VH domains of immunoglobulins. *J Mol Biol* 215(1):175-182, 1990.

Tuerk C and Gold L: Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* 249(4968):505-510, 1990.

van de Poll ML, Lafleur MV, van Gog F, Vrieling H, Meerman JH: N-acetylated and deacetylated 4'-fluoro-4-aminobiphenyl and 4-aminobiphenyl adducts differ in their ability to inhibit DNA replication of single-stranded M13 in vitro and of single-stranded phi X174 in Escherichia coli. *Carcinogenesis* 13(5):751-8, (May) 1992.

Vojtek AB, Hollenberg SM, Cooper JA: Mammalian Ras interacts directly with the serine/threonine kinase Raf. *Cell* 74(1):205-214, 1993.

Wenzler H, Mignery G, Fisher L, Park W: Sucrose-regulated expression of a chimeric potato tuber gene in leaves of transgenic tobacco plants. *Plant Mol Biol* 13(4):347-54, 1989.

Williams and Barclay, in Immunoglobulin Genes, The Immunoglobulin Gene Superfamily

Winnacker EL. From Genes to Clones: Introduction to Gene Technology. VCH Publishers, New York, NY, ©1987.

Winter G and Milstein C: Man-made antibodies. *Nature* 349(6307):293-299, 1991.

Yang X, Hubbard EJ, Carlson M: A protein kinase substrate identified by the two-hybrid system. *Science* 257(5070):680-2, (Jul 31) 1992.

USPN 4,683,195; Filed Feb. 7, 1986, Issued Jul 28. 1987. Mullis KB, Erlich HA, Arnheim N, Horn GT, Saiki RK, Scharf SJ: Process for Amplifying, Detecting, and/or Cloning Nucleic Acid Sequences.

USPN 4,683,202; Filed Oct. 25, 1985, Issued Jul. 28, 1987. Mullis KB: Process for Amplifying Nucleic Acid Sequences.

USPN 4,704,362; Filed Nov. 5, 1979, Issued Nov. 3, 1987. Itakura K, Riggs AD: Recombinant Cloning Vehicle Microbial Polypeptide Expression.

WO 88/08453; Filed Apr. 14, 1988, Published Nov. 3, 1988. Alakhov JB, Baranov, VI, Ovodov SJ, Ryabova LA, Spirin AS: Method of Obtaining Polypeptides in Cell-Free Translation System.

WO 90/05785; Filed Nov. 15, 1989, Published May 31, 1990. Schultz P: Method for Site-Specifically Incorporating Unnatural Amino Acids into Proteins.

WO 90/07003; Filed Jan. 27, 1989, Published June 28, 1990. Baranov VI, Morozov IJ, Spirin AS: Method for Preparative Expression of Genes in a Cell-free System of Conjugated Transcription/translation.

WO 91/02076; Filed June 14, 1990, Published Feb. 21, 1991. Baranov VI, Ryabova LA, Yarchuk OB, Spirin AS: Method for Obtaining Polypeptides in a Cell-free System.

WO 91/05058; Filed Oct. 5, 1989, Published Apr. 18, 1991. Kawasaki G: Cell-free Synthesis and Isolation of Novel Genes and Polypeptides.

WO 91/17271; Filed May 1, 1990, Published Nov. 14, 1991. Dower WJ, Cwirla SE: Recombinant Library Screening Methods.

WO 91/18980; Filed May 13, 1991, Published Dec. 12, 1991. Devlin JJ: Compositions and Methods for Indentifying Biologically Active Molecules.

WO 91/19818; Filed June 20, 1990, Published Dec. 26, 1991. Dower WJ, Cwirla SE, Barrett RW: Peptide Library and Screening Systems.

WO 92/02536; Filed Aug. 1, 1991, Published Feb. 20, 1992. Gold L, Tuerk C: Systematic Polypeptide Evolution by Reverse Translation.

WO 92/03918; Filed Aug. 28, 1991, Published Mar. 19, 1992. Lonberg N, Kay RM: Transgenic Non-human Animals Capable of Producing Heterologous Antibodies.

WO 92/03918; Filed Aug. 28, 1991, Published Mar. 19, 1992. Lonberg N, Kay RM: Transgenic Non-human Animals Capable of Producing Heterologous Antibodies.

WO 92/05258; Filed Sept. 17, 1991, Published Apr. 2, 1992. Fincher GB: Gene Encoding Barley Enzyme.

WO 92/14843; Filed Feb. 21, 1992, Published Sept. 3, 1992. Toole JJ, Griffin LC, Bock LC, Latham JA, Muenchau DD, Krawczyk S: Aptamers Specific for Biomolecules and Method of Making.

WO 93/08278; Filed Oct. 15, 1992, Published Apr. 29, 1993. Schatz PJ, Cull MG, Miller JF, Stemmer WP: Peptide Library and Screening Method.

WO 93/12227; Filed Dec 17, 1992, Published June 24, 1993. Lonberg, N; Kay RM: Transgenic Non-human Animals Capable of Producing Heterologous Antibodies.

WO 93/12227; Filed Dec. 17, 1992, Published June 24, 1993. Lonberg N, Kay RM: Transgenic Non-human Animals Capable of Producing Heterologous Antibodies.

WO 94/25585; Filed Apr. 25, 1994, Published Nov. 10, 1994. Lonberg, N, Kay RM: Transgenic Non-human Animals Capable of Producing Heterologous Antibodies.

WO 94/25585; Filed Apr. 25, 1994, Published Nov. 10, 1994. Lonberg N, Kay RM: Transgenic Non-human Animals Capable of Producing Heterologous Antibodies.

Arslan T, Abraham AT, Hecht SM: Structurally altered substrates for DNA topoisomerase I. Effects of inclusion of a single 3'-deoxynucleotide within the scissile strand. *Nucleosides Nucleotides* 1998 Jan-Mar;17(1-3):515-30.

Aupeix K, Toulme JJ: Binding of chemically-modified oligonucleotides to the double-stranded stem of an RNA hairpin. *Nucleosides Nucleotides* 1999 Jun-Jul;18(6-7):1647-50.

Bazzanini R, Manfredini S, Durini E, Groschel B, Cinatl J, Balzarini J, De Clercq E, Imbach JL, Perigaud C, Gosselin G: Prodrugs of Ara-CMP and Ara-AMP with a S-acyl-2-thioethyl (SATE) biolabile phosphate protecting group: synthesis and biological evaluation. *Nucleosides Nucleotides* 1999 Apr-May;18(4-5):971-2.

Blackburn GM, Liu X, Rosler A, Brenner C: Two hydrolase resistant analogues of diadenosine 5',5'''-P1,P3-triphosphate for studies with Fhit, the human fragile histidine triad protein. *Nucleosides Nucleotides* 1998 Jan-Mar;17(1-3):301-8.

Bridson PK, Lin X, Melman N, Ji XD, Jacobson KA: Synthesis and adenosine receptor affinity of 7-beta-D-ribofuranosylxanthine. *Nucleosides Nucleotides* 1998 Apr;17(4):759-68.

Brodin P, Gottikh M, Auclair C, Mouscadet JF: Inhibition of HIV-1 integration by mono- & bi-functionalized triple helix forming oligonucleotides. *Nucleosides Nucleotides* 1999 Jun-Jul;18(6-7):1717-8.

Creighton TE: Proteins Structures and Molecular Principles. New York: W. H. Freeman and Co., 1984.

De Clercq E: Carbocyclic adenosine analogues as S-adenosylhomocysteine hydrolase inhibitors and antiviral agents: recent advances. *Nucleosides Nucleotides* 1998 Jan-Mar;17(1-3):625-34.

de Zwart M, Link R, von Frijtag Drabbe Kunzel JK, Cristalli G, Jacobson KA, Townsend-Nicholson A, IJzerman AP: A functional screening of adenosine analogues at the adenosine A2B receptor: a search for potent agonists. *Nucleosides Nucleotides* 1998 Jun;17(6):969-85.

Egron D, Arzumanov AA, Dyatkina NB, Krayevsky A, Imbach JL, Aubertin AM, Gosselin G, Perigaud C: Synthesis, anti-HIV activity and stability studies of 3'-azido-2',3'-dideoxythymidine 5'-fluorophosphate. *Nucleosides Nucleotides* 1999 Apr-May;18(4-5):983-4

Gianolio DA, McLaughlin LW: Synthesis and triplex forming properties of pyrimidine derivative containing extended functionality. *Nucleosides Nucleotides* 1999 Aug;18(8):1751-69.

Gottikh MB, Volkov EM, Romanova EA, Oretskaya TS, Shabarova ZA: Synthesis of oligonucleotide-intercalator conjugates capable to inhibit HIV-1 DNA integration. *Nucleosides Nucleotides* 1999 Jun-Jul;18(6-7):1645-6.

Hotoda H, Koizumi M, Ohmine T, Furukawa H, Nishigaki T, Abe K, Kosaka T, Tsutsumi S, Sone J, Kaneko M: Biologically active oligodeoxyribonucleotides. 10: anti-HIV-1 activity and stability of modified hexanucleotides containing glycerol-skeleton. *Nucleosides Nucleotides* 1998 Jan-Mar;17(1-3):243-52.

JP10113194; Filed 19971022, Published 19980506. Donnelly, JJ ;Dwarki, VJ ;Liu, MA ;Montgomery, DL ;Parker, S ;Shiver, JW ;Ulmer JB: Nucleic Acid Preparation.

Kang SH, Sinhababu AK, Cho MJ: Synthesis and biological activity of bis(pivaloyloxymethyl) ester of 2'-azido-2'-deoxyuridine 5'-monophosphate. *Nucleosides Nucleotides* 1998 Jun;17(6):1089-98.

Krayevsky A, Arzumanov A, Shirokova E, Dyatkina N, Victorova L, Jasko M, Alexandrova L: dNTP modified at triphosphate residues: substrate properties towards DNA polymerases and stability in human serum. *Nucleosides Nucleotides* 1998 Jan-Mar;17(1-3):681-93.

Krayevsky AA, Dyatkina NB, Semizarov DG, Victorova LS, Shirokova EA, Theil F, Von Janta Lipinski MJ, Gosselin G, Imbach JL: Reasons and limits of substrate activity of modified L-dNTP in DNA biosynthesis. *Nucleosides Nucleotides* 1999 Apr-May;18(4-5):863-4.

Kvasyuk EI, Mikhailopulo IA, Suhadolnik RJ, Henderson EE, Muto NF, Iacono KT, Homon J, Pfleiderer W: Synthesis and biological activity of 2',5'-oligoadenylate trimers containing 5'-terminal 5'-amino-5'-deoxy- and 5'-amino-3',5'-dideoxyadenosine derivatives. *Nucleosides Nucleotides* 1999 Jun-Jul;18(6-7):1483-4.

Liu J, Skradis A, Kolar C, Kolath J, Anderson J, Lawson T, Talmadge J, Gmeiner WH: Increased cytotoxicity and decreased in vivo toxicity of FdUMP[10] relative to 5-FU. *Nucleosides Nucleotides* 1999 Aug;18(8):1789-802.

Lutz MJ, Will DW, Breipohl G, Benner SA, Uhlmann E: Synthesis of a monocharged peptide nucleic acid (PNA) analog and its recognition as substrate by DNA polymerases. *Nucleosides Nucleotides* 1999 Mar;18(3):393-401.

Monaco V, van de Wetering KI, Meeuwenoord NJ, van den Elst HA, Stuivenberg HR, Visse R, van der Kaaden JC, Moolenaar GF, Verhoeven EE, Goosen N, van der Marel GA, van Boom JH: Synthesis and biological evaluation of modified DNA fragments for the study of nucleotide excision repair in E. coli. *Nucleosides Nucleotides* 1999 Jun-Jul;18(6-7):1339-41.

Morozova OV, Kolpashchikov DM, Ivanova TM, Godovikova TS: Synthesis of new photocross-linking 5-C-base-substituted UTP analogs and their application in highly selective affinity labelling of the tick-borne encephalitis virus RNA replicase proteins. *Nucleosides Nucleotides* 1999 Jun-Jul;18(6-7):1513-4.

Nguyen-Ba N, Chan L, Quimpere M, Turcotte N, Lee N, Mitchell H, Bedard J: Design and SAR study of a novel class of nucleotide analogues as potent anti-HCMV agents. *Nucleosides Nucleotides* 1999 Apr-May;18(4-5):821-7.

Pandolfi D, Rauzi F, Capobianco ML: Evaluation of different types of end-capping modifications on the stability of oligonucleotides toward 3'- and 5'-exonucleases. *Nucleosides Nucleotides* 1999 Sep;18(9):2051-69.

Pankiewicz KW, Lesiak-Watanabe K: Novel mycophenolic adenine bis(phosphonate)s as potent anticancer agents and inducers of cells differentiation. *Nucleosides Nucleotides* 1999 Apr-May;18(4-5):927-32.

Perrin DM, Garestier T, Helene C: Expanding the catalytic repertoire of nucleic acid catalysts: simultaneous incorporation of two modified deoxyribonucleoside triphosphates bearing ammonium and imidazolyl functionalities. *Nucleosides Nucleotides* 1999 Mar;18(3):377-91.

Pfundheller HM, Koshkin AA, Olsen CE, Wengel J: Evaluation of oligonucleotides containing two novel 2'-O-methyl modified nucleotide monomers: a 3'-C-allyl and a 2'-O,3'-C-linked bicyclic derivative. *Nucleosides Nucleotides* 1999 Sep;18(9):2017-30.

Ramasamy KS, Stoisavljevic V: Synthesis and biophysical studies of modified oligonucleotides containing acyclic amino alcohol nucleoside analogs. *Nucleosides Nucleotides* 1999 Aug;18(8):1845-61.

Schinazi RF, Lesnikowski ZJ: Boron containing oligonucleotides. *Nucleosides Nucleotides* 1998 Jan-Mar;17(1-3):635-47.

Secrist JA 3rd, Parker WB, Allan PW, Bennett LL Jr, Waud WR, Truss JW, Fowler AT, Montgomery JA, Ealick SE, Wells AH, Gillespie GY, Gadi VK, Sorscher EJ: Gene therapy of cancer: activation of nucleoside prodrugs with E. coli purine nucleoside phosphorylase. *Nucleosides Nucleotides* 1999 Apr-May;18(4-5):745-57.

Shirokova EA, Shipitsin AV, Victorova LS, Dyatkina NB, Goryunova LE, Beabealashvilli RS, Hamilton CJ, Roberts SM, Krayevsky AA: Modified nucleoside 5'-triphosphonates as a new type of antiviral agents. *Nucleosides Nucleotides* 1999 Apr-May;18(4-5):1027-8.

Srivastava TK, Friedhoff P, Pingoud A, Katti SB: Application of oligonucleoside methylphosphonates in the studies on phosphodiester hydrolysis by Serratia endonuclease. *Nucleosides Nucleotides* 1999 Sep;18(9):1945-60.

Stattel JM, Yanachkov I, Wright GE: Synthesis and biochemical study of N2-(p-n-butylphenyl)-2'-deoxyguanosine 5'-(alpha,beta-imido)triphosphate (BuPdGMPNHPP): a non-substrate inhibitor of B family DNA polymerases. *Nucleosides Nucleotides* 1998 Aug;17(8):1505-13.

Terato H, Morita H, Ohyama Y, Ide H: Novel modification of 5-formyluracil by cysteine derivatives in aqueous solution. *Nucleosides Nucleotides* 1998 Jan-Mar;17(1-3):131-41.

Tomikawa A, Seno M, Sato-Kiyotaki K, Ohtsuki C, Hirai T, Yamaguchi T, Kawaguchi T, Yoshida S, Saneyoshi M: Synthetic nucleosides and nucleotides. 40. Selective inhibition of eukaryotic DNA polymerase alpha by 9-(beta-D-arabinofuranosyl)-2-(p-n-butylanilino) adenine 5'-triphosphate (BuAaraATP) and its 2'-up azido analog: synthesis and enzymatic evaluations. *Nucleosides Nucleotides* 1998 Jan-Mar;17(1-3):487-501.

US5580859; Filed 19940318, Issued 19961203. Felgner, PL. ;Wolff, JA. ;Rhodes, GH. ;Malone, RW. ;Carson, DA.: Delivery of exogenous DNA sequences in a mammal.

US5589466; Filed 19950126, Issued 19961231. Felgner, PL. ;Wolff, JA. ;Rhodes, GH. ;Malone, RW. ;Carson, DA.: Induction of a protective immune response in a mammal by injecting a DNA sequence.

US5641665; Filed 19941128, Issued 19970624. Hobart, PM. ;Margalith, M ;Parker, SE. ;Khatibi, S: Plasmids suitable for IL-2 expression.

US5693622; Filed 19950607, Issued 19971202. Wolff, JA. ;Duke, DJ. ;Felgner, PL.: Expression of exogenous polynucleotide sequences cardiac muscle of a mammal.

US5703055; Filed 19940126, Issued 19971230. Felgner, PL. ;Wolff, JA ;Rhodes, GH. ;Malone, RW ;Carson, DA.: Generation of antibodies through lipid mediated DNA delivery.

US5846946; Filed 19960614, Issued 19981208. Huebner, RC. ;Norman, JA. ;Liang, X ;Carner, KR. ;Barbour, AG. ;Luke, CJ.: Compositions and methods for administering Borrelia DNA.

US5910488; Filed 19951201, Issued 19990608. Nabel, GJ. ;Nabel, EG. ;Lew, D ;Marquet, M: Plasmids suitable for gene therapy.

Victorova LS, Semizarov DG, Shirokova EA, Alexandrova LA, Arzumanov AA, Jasko MV, Krayevsky AA: Human DNA polymerases and retroviral reverse transcriptases: selectivity in respect to dNTPs modified at triphosphate residues. *Nucleosides Nucleotides* 1999 Apr-May;18(4-5):1031-2.

von Janta-Lipinski M, Gaertner K, Lehmann C, Scheer H, Schildt J, Matthes E: Protein and RNA of human telomerase as targets for modified oligonucleotides. *Nucleosides Nucleotides* 1999 Jun-Jul;18(6-7):1719-20

WO9011092; Filed 19900321, A1 Published 19901004. Felgner, PL. ;Wolff, JA;Rhodes, GH. ;Malone, RW ;Carson, DA.: Expression Of Exogenus Polynucleotide Sequences In A Vertebrate.

WO9314778; Filed 19930121, A1 Published 19930805. Rhodes, GH. ;Dwarki, VJ.;Felgner, PL;Wang-Felgner, J;Manthorpe, M: Ex Vivo Gene Transfer.

WO9421797; Filed 19940314, A1 Published 19940929. Donnelly, JJ. ;Dwarki, VJ. ;Liu, MA. ;Montgomery, DL. ;Parker, SE. ;Shiver, JW. ;Ulmer, JB.: Nucleic Acid Pharmaceuticals.

WO9633736; Filed 19960426, A1 Published 19961031. Baruch DI; Pasloske BL; Howard, RJ: Malaria Peptides and Vaccines.

WO9735992; Filed 19970317, A1 Published 19971002. Hobart, PM. ;Liang, X: Tetracycline Inducible/Repressible Systems.

WO9926663; Filed 19981120, A2 Published 19990603. Horton, H ;Parker, S;Manthorpe, M; Felgner, P: Treatment Of Cancer Using Cytokine-Expressing Polynucleotides And Compositions Therefor.

WO9941368; Filed 19990210, A2 Published 19990819. Punnonen J, Stemmer WP, Whalen RG; Howard, R: Optimization of Immunomodulatory Properties of Genetic Vaccines.

WO9941369; Filed 19990210, A2 Published 19990819. Punnonen J, Stemmer WP, Whalen RG; Howard, R: Genetic Vaccine Vector Engineering.

WO9941383; Filed 19990210, A1 Published 19990819. Punnonen J, Bass, SH, Whalen, RG, Howard, R, Stemmer, WP: Antigen Library Immunization.

WO9941402; Filed 19990210, A2 Published 19990819. Punnonen J, Stemmer, WP, Howard R, Patten PA: Targeting of Genetic Vaccine Vectors.

3. CLAIMS

1. A method for obtaining an immunomodulatory polynucleotide that has an optimized modulatory effect on an immune response, or encodes a polypeptide that has an optimized modulatory effect on an immune response, the method comprising:

creating a library of non-stochastically generated progeny polynucleotides from a parental polynucleotide set;

wherein optimization can thus be achieved using one or more of the directed evolution methods as described herein in any combination, permutation and iterative manner;

whereby these directed evolution methods include the introduction of mutations by non-stochastic methods, including by "gene site saturation mutagenesis" as described herein;

and whereby these directed evolution methods also include the introduction mutations by non-stochastic polynucleotide reassembly methods as described herein; including by synthetic ligation polynucleotide reassembly as described herein.

2. The method of claim 1, wherein said optimized modulatory effect on an immune response is induced by a genetic vaccine vector.

3. A method for obtaining an immunomodulatory polynucleotide that has an optimized modulatory effect on an immune response, or encodes a polypeptide that has an optimized modulatory effect on an immune response, the method comprising:

screening a library of non-stochastically generated progeny polynucleotides to identify an optimized non-stochastically generated progeny polynucleotide that has, or

encodes a polypeptide that has, a modulatory effect on an immune response; wherein the optimized non-stochastically generated polynucleotide or the polypeptide encoded by the non-stochastically generated polynucleotide exhibits an enhanced ability to modulate an immune response compared to a parental polynucleotide from which the library was created.

4. The method of claim 3, wherein said optimized modulatory effect on an immune response is induced by a genetic vaccine vector.

5. A method for obtaining an immunomodulatory polynucleotide that has an optimized modulatory effect on an immune response, or encodes a polypeptide that has an optimized modulatory effect on an immune response, the method comprising:

a) creating a library of non-stochastically generated progeny polynucleotides from a parental polynucleotide set; and

b) screening the library to identify an optimized non-stochastically generated progeny polynucleotide that has, or encodes a polypeptide that has, a modulatory effect on an immune response induced by a genetic vaccine vector; wherein the optimized non-stochastically generated polynucleotide or the polypeptide encoded by the non-stochastically generated polynucleotide exhibits an enhanced ability to modulate an immune response compared to a parental polynucleotide from which the library was created;

whereby optimization can thus be achieved using one or more of the directed evolution methods as described herein in any combination, permutation, and iterative manner;

whereby these directed evolution methods include the introduction of point mutations by non-stochastic methods, including by "gene site saturation mutagenesis" as described herein;

and whereby these directed evolution methods also include the introduction mutations by non-stochastic polynucleotide reassembly methods as described herein; including by synthetic ligation polynucleotide reassembly as described herein.

6. The method of claim 5, wherein said optimized modulatory effect on an immune response is induced by a genetic vaccine vector.

7. The method of any of claims 1-6, wherein the optimized non-stochastically generated polynucleotide is incorporated into a genetic vaccine vector.

8. The method of any of claims 1-6, wherein the optimized non-stochastically generated polynucleotide, or a polypeptide encoded by the optimized non-stochastically generated polynucleotide, is administered in conjunction with a genetic vaccine vector.

9. The method of any of claims 1-6, wherein the library of non-stochastically generated progeny polynucleotides is created by a process selected from the group consisting of gene reassembly, oligonucleotide-directed saturation mutagenesis, and any combination, permutation and iterative manner.

10. The method of any of claims 1-6, wherein the optimized non-stochastically generated polynucleotide that has a modulatory effect on an immune response is obtained by:

a) non-stochastically reassembling at least two parental template polynucleotide, each of which is, or encodes a molecule that is, involved in modulating an immune response;

wherein the first and second parental templates differ from each other in two or more nucleotides, to produce a library of non-stochastically generated polynucleotides; and

b) screening the library to identify at least one optimized non-stochastically generated polynucleotide that exhibits, either by itself or through the encoded molecule, an enhanced ability to modulate an immune response in comparison to a parental polynucleotide from which the library was created.

11. The method of claim 10, wherein the method further comprises the steps of:

c) subjecting a working optimized non-stochastically generated polynucleotide to a further round of non-stochastic reassembly with at least one additional polynucleotide, which is the same or different from the first and second polynucleotides, to produce a further working library of recombinant polynucleotides;

d) screening the further working library to identify at least one further optimized non-stochastically generated polynucleotide that exhibits an enhanced ability to modulate an immune response in comparison to a parental polynucleotide from which the library was created; and

e) optionally repeating c) and d) as necessary, until a desirable further optimized non-stochastically generated polynucleotide that exhibits an enhanced ability to modulate an immune response than a form of the nucleic acid from which the library was created.

12. The method of any of claims 1-6, wherein the optimized non-stochastically generated polynucleotide encodes a polypeptide that can interact with a cellular receptor involved in mediating an immune response; wherein the polypeptide acts as an agonist or antagonist of the receptor.

13. The method of claim 12, wherein the cellular receptor is a macrophage scavenger receptor.

14. The method of claim 12, wherein the cellular receptor is selected from the group consisting of a cytokine receptor and a chemokine receptor.

15. The method of claim 14, wherein the chemokine receptor is CCR6.

16. The method of claim 12, wherein the polypeptide mimics the activity of a natural ligand for the receptor but does not induce immune reactivity to said natural ligand.

17. The method of claim 12, wherein the library is screened by:

i) expressing the non-stochastically generated progeny polynucleotides so that the encoded polypeptides are produced as fusions with a protein displayed on the surface of a replicable genetic package;

ii) contacting the replicable genetic packages with a plurality of cells that display the receptor; and

iii) identifying cells that exhibit a modulation of an immune response mediated by the receptor.

18. The method of claim 17, wherein the replicable genetic package is selected from the group consisting of a bacteriophage, a cell, a spore, and a virus.

19. The method of claim 18, wherein the replicable genetic package is an M13 bacteriophage and the protein is encoded by geneIII or geneVIII.

20. The method of claim 12, which method further comprises introducing the optimized non-stochastically generated polynucleotide into a genetic vaccine vector and administering the vector to a mammal, wherein the peptide or polypeptide is expressed and acts as an agonist or antagonist of the receptor.

21. The method of claim 12, which method further comprises producing the polypeptide encoded by the optimized non-stochastically generated polynucleotide and introducing the polypeptide into a mammal in conjunction with a genetic vaccine vector.

22. The method of claim 12, wherein the optimized non-stochastically generated polynucleotide is inserted into an antigen-encoding nucleotide sequence of a genetic vaccine vector.

23. The method of claim 22, wherein the optimized non-stochastically generated polypeptide is introduced into a nucleotide sequence that encodes an M- loop of an HBsAg polypeptide.

24. The method of any of claims 1-6, wherein the optimized non-stochastically generated polynucleotide comprises a nucleotide sequence rich in unmethylated CpG.

25. The method of any of claims 1-6, wherein the optimized non-stochastically generated polynucleotide encodes a polypeptide that inhibits an allergic reaction.

26. The method of claim 25, wherein the polypeptide is selected from the group consisting of interferon- , interferon- , IL- 10, IL- 12, an antagonist of IL-4, an antagonist of IL-5, and an antagonist of IL-13.

27. The method of 1, wherein the optimized recombinant polynucleotide encodes an antagonist of IL-10.

28. The method of claim 27, wherein the antagonist of IL-10 is soluble or defective IL-10 receptor or IL-20/MDA-7.

29. The method of any of claims 1-6, wherein the optimized non-stochastically generated polynucleotide encodes a co-stimulator.

30. The method of claim 29, wherein the co-stimulator is B7-1 (CD80) or B7-2 (CD86) and the screening step involves selecting variants with altered activity through CD28 or CTLA-4.

31. The method of claim 29, wherein the co-stimulator is CD1, CD40, CD154 (ligand for CD40) or CD150 (SLAM).

32. The method of claim 29, wherein the co-stimulator is a cytokine.

33. The method of claim 32, wherein the cytokine is selected from the group consisting of IL-1, IL-2, IL-3, IL-4, IL-5, IL-6, IL-7, IL-8, IL-9, IL- 10, IL- 11, IL- 12, IL-13, IL-14, IL-15, IL-16, IL-17, IL-18, GM-CSF, G-CSF, TNF- , IFN- , IFN- , and IL-20 (MDA-7).

34. The method of 33, wherein the library of non-stochastically generated polynucleotides is screened by testing the ability of cytokines encoded by the non-stochastically generated polynucleotides to activate cells which contain a receptor for the cytokine.

35. The method of claim 34, wherein the cells contain a heterologous nucleic acid that encodes the receptor for the cytokine.

36. The method of 33, wherein the cytokine is interleukin-12 and the screening is performed by: growing mammalian cells which contain the genetic vaccine vector in a culture medium; and detecting whether T cell proliferation or T cell differentiation is induced by contact with the culture medium.

37. The method of 33, wherein the cytokine is interferon- and the screening is performed by:

i) expressing the non-stochastically generated polynucleotides so that the encoded polypeptides are produced as fusions with a protein displayed on the surface of a replicable genetic package;

ii) contacting the replicable genetic packages with a plurality of B cells; and

iii) identifying phage library members that are capable of inhibiting proliferation of the B cells.

38. The method of claim 33, wherein the immune response of interest is differentiation of T cells to T_H1 cells and the screening is performed by contacting a population of T cells with the cytokines encoded by the members of the library of recombinant polynucleotides and identifying library members that encode a cytokine that induces the T cells to produce IL-2 and interferon- .

39. The method of claim 32, wherein the cytokine encoded by the optimized non-stochastically generated polynucleotide exhibits reduced immunogenicity compared to a

cytokine encoded by a non-optimized polynucleotide, and the reduced immunogenicity is detected by introducing a cytokine encoded by the non-stochastically generated polynucleotide into a mammal and determining whether an immune response is induced against the cytokine.

40. The method of claim 29, wherein the co-stimulator is B7-1 (CD80) or B7-2 (CD86) and the cell is tested for ability to costimulate an immune response.

41. The method of any of claims 1-6, wherein the optimized recombinant polynucleotide encodes a cytokine antagonist.

42. The method of claim 41, wherein the cytokine antagonist is selected from the group consisting of a soluble cytokine receptor and a transmembrane cytokine receptor having a defective signal sequence.

43. The method of claim 41, wherein the cytokine antagonist is selected from the group consisting of IL- 1 OR and IL-4R.

44. The method of any of claims 1-6, wherein the optimized non-stochastically generated polynucleotide encodes a polypeptide capable of inducing a predominantly T_H1 immune response.

45. The method of any of claims 1-6, wherein the optimized non-stochastically generated polynucleotide encodes a polypeptide capable of inducing a predominantly T_H2 immune response.

46. The method of any of claims 1-6, wherein said optimized modulatory effect on an immune response is a decrease in an unwanted modulatory effect on an immune response;

whereby application of the method can be used to generate a molecule having a decreased ability to elicit an immune response from a host recipient of said molecule, where said recipient can be a human or an animal host;

and whereby application of the method can thus be used to generate a molecule having decreased antigenicity with respect to at least one host recipient of said molecule.

47. The method of any of claims 1-6, wherein said optimized modulatory effect on an immune response is an increase in a desirable modulatory effect on an immune response;

whereby application of the method can be used to generate a molecule having an increased ability to elicit an immune response from a host recipient of said molecule, where said recipient can be a human or an animal host;

and whereby application of the method can thus be used to generate a molecule having increased antigenicity with respect to at least one host recipient of said molecule.

48. The method of any of claims 1-6, wherein said optimized modulatory effect on an immune response is both a decrease in a first unwanted modulatory effect on an immune response as well as an increase in a second desirable modulatory effect on an immune response;

whereby application of the method can be used to generate a molecule having both a decreased ability to elicit a first immune response from a first host recipient of said molecule as well as a an increased ability to elicit a second immune response from a second host recipient of said molecule;

whereby the first and the second recipient hosts can be the same or different;

whereby each of the first and the second recipient hosts can be a human or an animal host;

and whereby application of the method can thus be used to generate a molecule having both a first decreased antigenicity with respect to at least one host recipient of said molecule and a second decreased antigenicity with respect to at least one host recipient of said molecule.

49. The method of claim 48, wherein said first and said second modulatory effect on an immune response are evolved for respectively a first and a second module on the same multimodule vaccine vector;

whereby a module is exemplified by the following modules, as well as by a fragment derivative or analog thereof: an antigen coding sequence, a polyadenylation sequence, a sequence coding for a co-stimulatory molecule, a sequence coding for an inducible repressor or transactivator, a eukaryotic origin or replication, a prokaryotic origin of replication, a sequence coding for a prokaryotic marker, , and enhancer, a promoter, and operator, and an intron.

50. The method of any of claims 1-6, wherein the optimized modulatory effect on an immune response is comprised of an increase in the stability of the immunomodulatory (IM) polynucleotide or polypeptide encoded thereby;

whereby application of the method can be used to generate a molecule having an increased stability ex vivo, thus, for example, increasing shelf-life and/or ease of storage and/or length of time before expiration of activity upon storage;

and whereby application of the method can also be used to generate a molecule having an increased stability in vivo upon administration to a host recipient, thus, for example, increasing resistance to digestive acids and/or increasing stability in the circulation and/or any other method of elimination or destruction by the host recipient.

51. The method of any of claims 1-6, wherein the immunomodulatory (IM) polynucleotide or polypeptide encoded thereby; has an optimized modulatory effect on an immune response in a human host recipient;

whereby application of the method can thus be used to generate an optimized genetic vaccine for human recipeints.

52. The method of any of claims 1-6, wherein the immunomodulatory (IM) polynucleotide or polypeptide encoded thereby; has an optimized modulatory effect on an immune response in an animal host recipient;

whereby application of the method can thus be used to generate an optimized genetic vaccine for animal recipients, including animals that are farmed or raised by man, animals that are not farmed or raised by man, domesticated animals, and non-domesticated animals.

53. A method for obtaining an optimized polynucleotide that encodes an accessory molecule that improves the transport or presentation of antigens by a cell, the method comprising:

a) creating a library of non-stochastically generated polynucleotides by subjecting to optimization by non-stochastic directed evolution a parental polynucleotide set in which is encoded all or part of the accessory molecule; and

b) screening the library to identify an optimized non-stochastically generated progeny polynucleotide that encodes a recombinant molecule that confers upon a cell an increased or decreased ability to transport or present an antigen on a surface of the cell compared to an accessory molecule encoded by template polynucleotides not subjected to the non-stochastic reassembly;

whereby application of the method can thus be used to generate an optimized molecule for human recipients &/or animal recipients, including animals that are farmed or raised by man, animals that are not farmed or raised by man, domesticated animals, and non-domesticated animals;

whereby optimization can thus be achieved using one or more of the directed evolution methods as described herein in any combination, permutation, and iterative manner;

whereby these directed evolution methods include the introduction of point mutations by non-stochastic methods, including by "gene site saturation mutagenesis" as described herein;

and whereby these directed evolution methods also include the introduction mutations by non-stochastic polynucleotide reassembly methods as described herein; including by synthetic ligation polynucleotide reassembly as described herein.

54. The method of claim 53, wherein the screening involves:

i) introducing the library of non-stochastically generated polynucleotides into a genetic vaccine vector that encodes an antigen to form a library of vectors; introducing the library of vectors into mammalian cells; and

ii) identifying mammalian cells that exhibit increased or decreased immunogenicity to the antigen.

55. The method of claim 53, wherein the accessory molecule comprises a proteasome or a TAP polypeptide.

56. The method of claim 53, wherein the accessory molecule comprises a cytotoxic T-cell inducing sequence.

57. The method of claim 56, wherein the cytotoxic T-cell inducing sequence is obtained from a hepatitis B surface antigen.

58. The method of claim 53, wherein the accessory molecule comprises an immunogenic agonist sequence.

59. A method for obtaining an immunomodulatory polynucleotide that has, an optimized expression in a recombinant expression host, the method comprising:

creating a library of non-stochastically generated progeny polynucleotides from a parental polynucleotide set;

whereby optimization can thus be achieved using one or more of the directed evolution methods as described herein in any combination, permutation and iterative manner;

whereby these directed evolution methods include the introduction of mutations by non-stochastic methods, including by "gene site saturation mutagenesis" as described herein;

and whereby these directed evolution methods also include the introduction mutations by non-stochastic polynucleotide reassembly methods as described herein; including by synthetic ligation polynucleotide reassembly as described herein.

60. A method for obtaining an immunomodulatory polynucleotide that has an optimized expression in a recombinant expression host, the method comprising:

screening a library of non-stochastically generated progeny polynucleotides to identify an optimized non-stochastically generated progeny polynucleotide that has an optimized expression in a recombinant expression host when compared to the expression of a parental polynucleotide from which the library was created.

61. A method for obtaining an immunomodulatory polynucleotide that has an optimized expression in a recombinant expression host, the method comprising:

a) creating a library of non-stochastically generated progeny polynucleotides from a parental polynucleotide set; and

b) screening a library of non-stochastically generated progeny polynucleotides to identify an optimized non-stochastically generated progeny polynucleotide that has an optimized expression in a recombinant expression host when compared to the expression of a parental polynucleotide from which the library was created;

whereby optimization can thus be achieved using one or more of the directed evolution methods as described herein in any combination, permutation, and iterative manner;

whereby these directed evolution methods include the introduction of point mutations by non-stochastic methods, including by "gene site saturation mutagenesis" as described herein;

and whereby these directed evolution methods also include the introduction mutations by non-stochastic polynucleotide reassembly methods as described herein; including by synthetic ligation polynucleotide reassembly as described herein.

62. The method of any of claims 59-61, wherein the recombinant expression host is a prokaryote.

63. The method of any of claims 59-61, wherein the recombinant expression host is a eukaryote.

64. The method of claim 63, wherein the recombinant expression host is a plant.

65. The method of any of claims 64, wherein the recombinant expression host is a monocot.

66. The method of any of claims 64, wherein the recombinant expression host is a dicot.

67. The method of any of claims 1-6, 53, or 59-61, wherein creating a library of non-stochastically generated progeny polynucleotides from a parental polynucleotide set is comprised of subjecting the parental polynucleotide set to "gene site saturation mutagenesis" as described herein.

68. The method of any of claims 1-6, 53, or 59-61, wherein creating a library of non-stochastically generated progeny polynucleotides from a parental polynucleotide set is comprised of subjecting the parental polynucleotide set to "synthetic ligation polynucleotide reassembly" as described herein.

69. The method of any of claims 1-6, 53, or 59-61, wherein creating a library of non-stochastically generated progeny polynucleotides from a parental polynucleotide set is comprised of subjecting the parental polynucleotide set to both "gene site saturation mutagenesis" as described herein, and to "synthetic ligation polynucleotide reassembly" as described herein.

70. A method of producing a progeny polynucleotide set by subjecting a double-stranded circular parental polynucleotide molecule to mutagenesis, said method comprising the steps of:

a) annealing a first primer and a second primer to said parental polynucleotide molecule;

wherein said first primer is comprised of a first primer sequence that is complementary to a first annealment region of the parental polynucleotide molecule,

wherein said second primer is comprised of a second primer sequence that is complementary to a second annealment region of the parental polynucleotide molecule,

wherein said first annealment region and said second annealment region are non-overlapping and therefore staggered,

and wherein at least one of said first and second primers contains a non-stochastic mutagenic cassette with respect to the parental polynucleotide molecule; and

b) synthesizing by means of a polymerase-catalyzed amplification reaction a first progeny polynucleotide strand comprised of said first primer and a second progeny polynucleotide strand comprised of said second primer;

wherein the first progeny polynucleotide strand and the second progeny polynucleotide strand may form a double-stranded mutagenized circular polynucleotide product.

71. A method of producing a progeny polynucleotide set by subjecting a double-stranded circular parental polynucleotide molecule to mutagenesis, said method comprising the steps of:

a) annealing a first primer and a second primer to said parental polynucleotide molecule;

wherein said first primer is comprised of a first primer sequence that is complementary to a first annealment region of the parental polynucleotide molecule,

wherein said second primer is comprised of a second primer sequence that is complementary to a second annealment region of the parental polynucleotide molecule,

wherein said first annealment region and said second annealment region are non-overlapping and therefore staggered,

wherein at least one of said first and second primers contains a non-stochastic mutagenic cassette with respect to the parental polynucleotide molecule, and

wherein said non-stochastic mutagenic cassette contained in said at least one primer is degenerate in nature; and

b) synthesizing by means of a polymerase-catalyzed amplification reaction a first progeny polynucleotide strand comprised of said first primer and a second progeny polynucleotide strand comprised of said second primer;

wherein the first progeny polynucleotide strand and the second progeny polynucleotide strand may form a double-stranded mutagenized circular polynucleotide product;

whereby the generation of a degenerate progeny polynucleotide set may be achieved by applying said method.

72. A method for producing from a template polypeptide a set of progeny polypeptides in which a non-stochastic range of single amino acid substitutions is represented at each amino acid position, comprising the steps of:

a) subjecting a codon-containing template polynucleotide to polymerase-based amplification using a degenerate oligonucleotide for each codon to be mutagenized, wherein each of said degenerate oligonucleotides is comprised of a

first homologous sequence and a degenerate trinucleotide cassette, so as to generate a set of progeny polynucleotides; and

b) subjecting said set of progeny polynucleotides to clonal amplification such that polypeptides encoded by the progeny polynucleotides are expressed;

whereby, said method provides a means for generating a predetermined number of amino acids to be represented at each amino acid site along a parental polypeptide template, up to as many as all 20 amino acids at each of said amino acid sites.

73. The method of claim 72, wherein said degenerate oligonucleotide is comprised of a first homologous sequence, a degenerate trinucleotide cassette, and a second homologous sequence.

74. The method of claim 72, wherein said degenerate trinucleotide cassette is comprised of a first mononucleotide cassette selected from the group consisting of:

a degenerate A/C mononucleotide cassette,
a degenerate A/G mononucleotide cassette,
a degenerate A/T mononucleotide cassette,
a degenerate C/G mononucleotide cassette,
a degenerate C/T mononucleotide cassette,
a degenerate G/T mononucleotide cassette,
a degenerate C/G/T mononucleotide cassette,
a degenerate A/G/T mononucleotide cassette,
a degenerate A/C/T mononucleotide cassette,
a degenerate A/C/G mononucleotide cassette,
and a degenerate N or A/C/G/T mononucleotide cassette;

and wherein said degenerate trinucleotide cassette is further comprised of a second and a third mononucleotide cassette, each selected from the group consisting of:

a degenerate A/C mononucleotide cassette,

a degenerate A/G mononucleotide cassette,

a degenerate A/T mononucleotide cassette,

a degenerate C/G mononucleotide cassette,

a degenerate C/T mononucleotide cassette,

a degenerate G/T mononucleotide cassette,

a degenerate C/G/T mononucleotide cassette

a degenerate A/G/T mononucleotide cassette,

a degenerate A/C/T mononucleotide cassette,

a degenerate A/C/G mononucleotide cassette,

a degenerate N or A/C/G/T mononucleotide cassette,

a non-degenerate A mononucleotide cassette,

a non-degenerate C mononucleotide cassette,

a non-degenerate G mononucleotide cassette,

and a non-degenerate T mononucleotide cassette.

75. The method of claim 72, where said degenerate trinucleotide cassette is selected from the group consisting of:

a degenerate N,N,N trinucleotide cassette,

a degenerate N,N,G/T trinucleotide cassette,

a degenerate N,N,G/C trinucleotide cassette,

a degenerate N,N,A/C/G trinucleotide cassette,

a degenerate N,N,A/G/T trinucleotide cassette,

and a degenerate N,N,C/G/T trinucleotide cassette;

whereby, said method provides a means for generating all 20 amino acid changes at each amino acid site along a parental polypeptide template, because the degeneracy of the specified trinucleotide cassette sequences includes codons for all 20 amino acids.

76. The method of claim 72, wherein said degenerate oligonucleotide is comprised of a first homologous sequence and a plurality of trinucleotide cassettes;

whereby, said method provides a means for generating a progeny polypeptide having a plurality of concurrent single amino acid changes with respect to a parental polypeptide template.

77. The method of claim 76, wherein each of said degenerate trinucleotide cassettes is comprised of a first mononucleotide cassette selected from the group consisting of:

a degenerate A/C mononucleotide cassette,
a degenerate A/G mononucleotide cassette,
a degenerate A/T mononucleotide cassette,
a degenerate C/G mononucleotide cassette,
a degenerate C/T mononucleotide cassette,
a degenerate G/T mononucleotide cassette,
a degenerate C/G/T mononucleotide cassette,
a degenerate A/G/T mononucleotide cassette,
a degenerate A/C/T mononucleotide cassette,
a degenerate A/C/G mononucleotide cassette,

and a degenerate N or A/C/G/T mononucleotide cassette;

and wherein each of said degenerate trinucleotide cassettes is further comprised of a second and a third mononucleotide cassette, each selected from the group of consisting of:

a degenerate A/C mononucleotide cassette,
a degenerate A/G mononucleotide cassette,
a degenerate A/T mononucleotide cassette,
a degenerate C/G mononucleotide cassette,
a degenerate C/T mononucleotide cassette,
a degenerate G/T mononucleotide cassette,
a degenerate C/G/T mononucleotide cassette
a degenerate A/G/T mononucleotide cassette,
a degenerate A/C/T mononucleotide cassette,
a degenerate A/C/G mononucleotide cassette,
a degenerate N or A/C/G/T mononucleotide cassette,
a non-degenerate A mononucleotide cassette,
a non-degenerate C mononucleotide cassette,
a non-degenerate G mononucleotide cassette,
and a non-degenerate T mononucleotide cassette.

78. The method of claim 76, where said degenerate trinucleotide cassette is selected from the group consisting of:

a degenerate N,N,N trinucleotide cassette,
a degenerate N,N,G/T trinucleotide cassette,
a degenerate N,N,G/C trinucleotide cassette,
a degenerate N,N,A/C/G trinucleotide cassette,

a degenerate N,N,A/G/T trinucleotide cassette,
and a degenerate N,N,C/G/T trinucleotide cassette;

whereby, said method provides a means for generating all 20 amino acid changes at each amino acid site along a parental polypeptide template, because the degeneracy of the specified trinucleotide cassette sequences includes codons for all 20 amino acids.

79. The method of claim 72, wherein said degenerate oligonucleotide is comprised of a first homologous sequence, and a plurality of trinucleotide cassettes, and a second homologous sequence.

80. A method for producing from a template polypeptide a set of progeny polypeptides in which a non-stochastic range of single amino acid substitutions is represented at each amino acid position, and for identifying desirable amino acid substitutions and combinations thereof among the progeny molecules, comprising the steps of:

a) subjecting a codon-containing template polynucleotide to polymerase-based amplification using a degenerate oligonucleotide cassette for each codon to be mutagenized, wherein each of said degenerate oligonucleotides is comprised of a first homologous sequence and a degenerate trinucleotide cassette, so as to generate a set of progeny polynucleotides; and

b) subjecting said set of progeny polynucleotides to clonal amplification such that polypeptides encoded by the progeny polynucleotides are expressed; and

c) subjecting said expressed progeny polypeptides to screening in order to compare them to the parental polynucleotide with respect to at least one molecular property of interest;

whereby, said method provides a means for generating a predetermined number of amino acids to be represented at each amino acid site along a parental polypeptide template, up to as many as all 20 amino acids at each of said amino acid sites; and

whereby, said method provides a means for identifying among said progeny polypeptides those that display a desirable change with respect to at least one molecular property when compared with its parental polypeptide.

81. The method of claim 80, wherein said degenerate trinucleotide cassette is comprised of a first nucleotide selected from the group consisting of:

a degenerate A/C mononucleotide cassette,
a degenerate A/G mononucleotide cassette,
a degenerate A/T mononucleotide cassette,
a degenerate C/G mononucleotide cassette,
a degenerate C/T mononucleotide cassette,
a degenerate G/T mononucleotide cassette,
a degenerate C/G/T mononucleotide cassette,
a degenerate A/G/T mononucleotide cassette,
a degenerate A/C/T mononucleotide cassette,
a degenerate A/C/G mononucleotide cassette,
and a degenerate N or A/C/G/T mononucleotide cassette;

and wherein said degenerate trinucleotide cassette is further comprised of a second and a third mononucleotide cassette, each selected from the group consisting of:

a degenerate A/C mononucleotide cassette,
a degenerate A/G mononucleotide cassette,
a degenerate A/T mononucleotide cassette,
a degenerate C/G mononucleotide cassette,
a degenerate C/T mononucleotide cassette,
a degenerate G/T mononucleotide cassette,
a degenerate C/G/T mononucleotide cassette
a degenerate A/G/T mononucleotide cassette,
a degenerate A/C/T mononucleotide cassette,
a degenerate A/C/G mononucleotide cassette,
a degenerate N or A/C/G/T mononucleotide cassette,
a non-degenerate A mononucleotide cassette,
a non-degenerate C mononucleotide cassette,
a non-degenerate G mononucleotide cassette,
and a non-degenerate T mononucleotide cassette.

82. The method of claim 80, where said degenerate trinucleotide cassette is selected from the group consisting of:

a degenerate N,N,N trinucleotide cassette,
a degenerate N,N,G/T trinucleotide cassette,
a degenerate N,N,G/C trinucleotide cassette,
a degenerate N,N,A/C/G trinucleotide cassette,
a degenerate N,N,A/G/T trinucleotide cassette,
and a degenerate N,N,C/G/T trinucleotide cassette;

whereby, said method provides a means for generating all 20 amino acid changes at each amino acid site along a parental polypeptide template, because the

degeneracy of the specified trinucleotide cassette sequences includes codons for all 20 amino acids.

83. The method of claim 80, wherein said degenerate oligonucleotide is comprised of a first homologous sequence and a plurality of trinucleotide cassettes;

whereby, said method provides a means for generating a progeny polypeptide having a plurality of concurrent single amino acid changes with respect to a parental polypeptide template.

84. The method of claim 80, wherein each of said degenerate trinucleotide cassettes is comprised of a first mononucleotide cassette selected from the group consisting of:

a degenerate A/C mononucleotide cassette,
a degenerate A/G mononucleotide cassette,
a degenerate A/T mononucleotide cassette,
a degenerate C/G mononucleotide cassette,
a degenerate C/T mononucleotide cassette,
a degenerate G/T mononucleotide cassette,
a degenerate C/G/T mononucleotide cassette,
a degenerate A/G/T mononucleotide cassette,
a degenerate A/C/T mononucleotide cassette,
a degenerate A/C/G mononucleotide cassette,
and a degenerate N or A/C/G/T mononucleotide cassette;

and wherein each of said degenerate trinucleotide cassettes is further comprised of a second and a third mononucleotide cassette, each selected from the group consisting of:

a degenerate A/C mononucleotide cassette,

a degenerate A/G mononucleotide cassette,

a degenerate A/T mononucleotide cassette,

a degenerate C/G mononucleotide cassette,

a degenerate C/T mononucleotide cassette,

a degenerate G/T mononucleotide cassette,

a degenerate C/G/T mononucleotide cassette

a degenerate A/G/T mononucleotide cassette,

a degenerate A/C/T mononucleotide cassette,

a degenerate A/C/G mononucleotide cassette,

a degenerate N or A/C/G/T mononucleotide cassette,

a non-degenerate A mononucleotide cassette,

a non-degenerate C mononucleotide cassette,

a non-degenerate G mononucleotide cassette,

and a non-degenerate T mononucleotide cassette.

85. The method of claim 80, where said degenerate trinucleotide cassette is selected from the group consisting of:

a degenerate N,N,N trinucleotide cassette,
a degenerate N,N,G/T trinucleotide cassette,
a degenerate N,N,G/C trinucleotide cassette,
a degenerate N,N,A/C/G trinucleotide cassette,
a degenerate N,N,A/G/T trinucleotide cassette,
and a degenerate N,N,C/G/T trinucleotide cassette;

whereby, said method provides a means for generating all 20 amino acid changes at each amino acid site along a parental polypeptide template, because the degeneracy of the specified trinucleotide cassette sequences includes codons for all 20 amino acids.

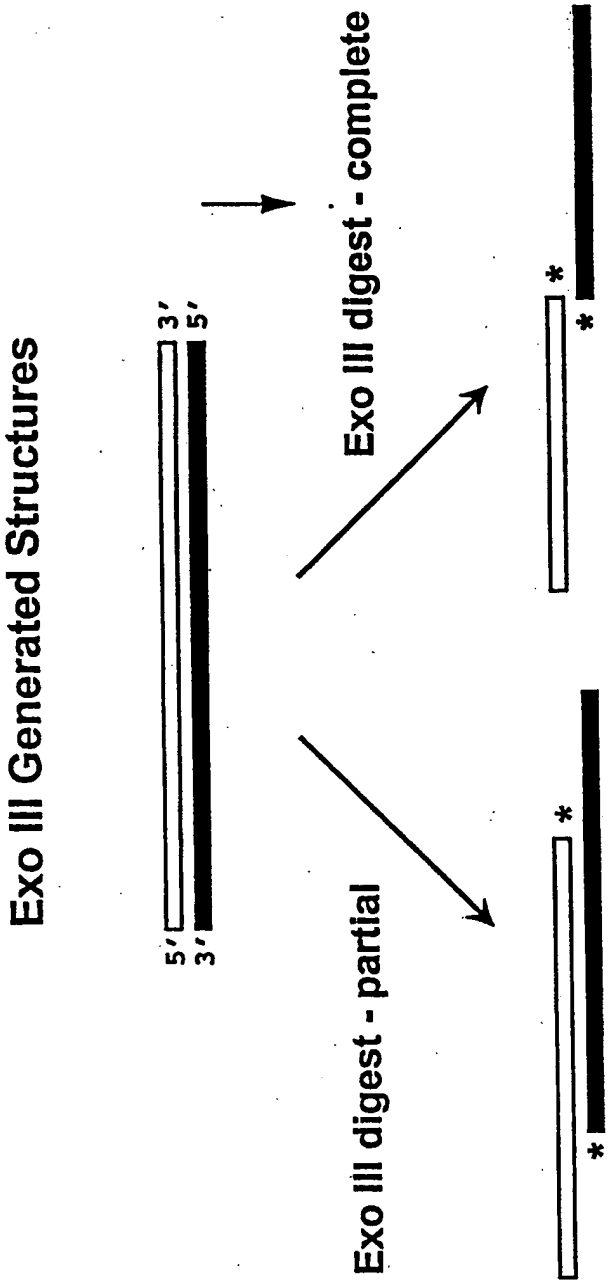

Exo III Generated Structures
3'
5'
5'
3'
Exo III digest - complete
Exo III digest - partial
*
*
*
*

Figure 1

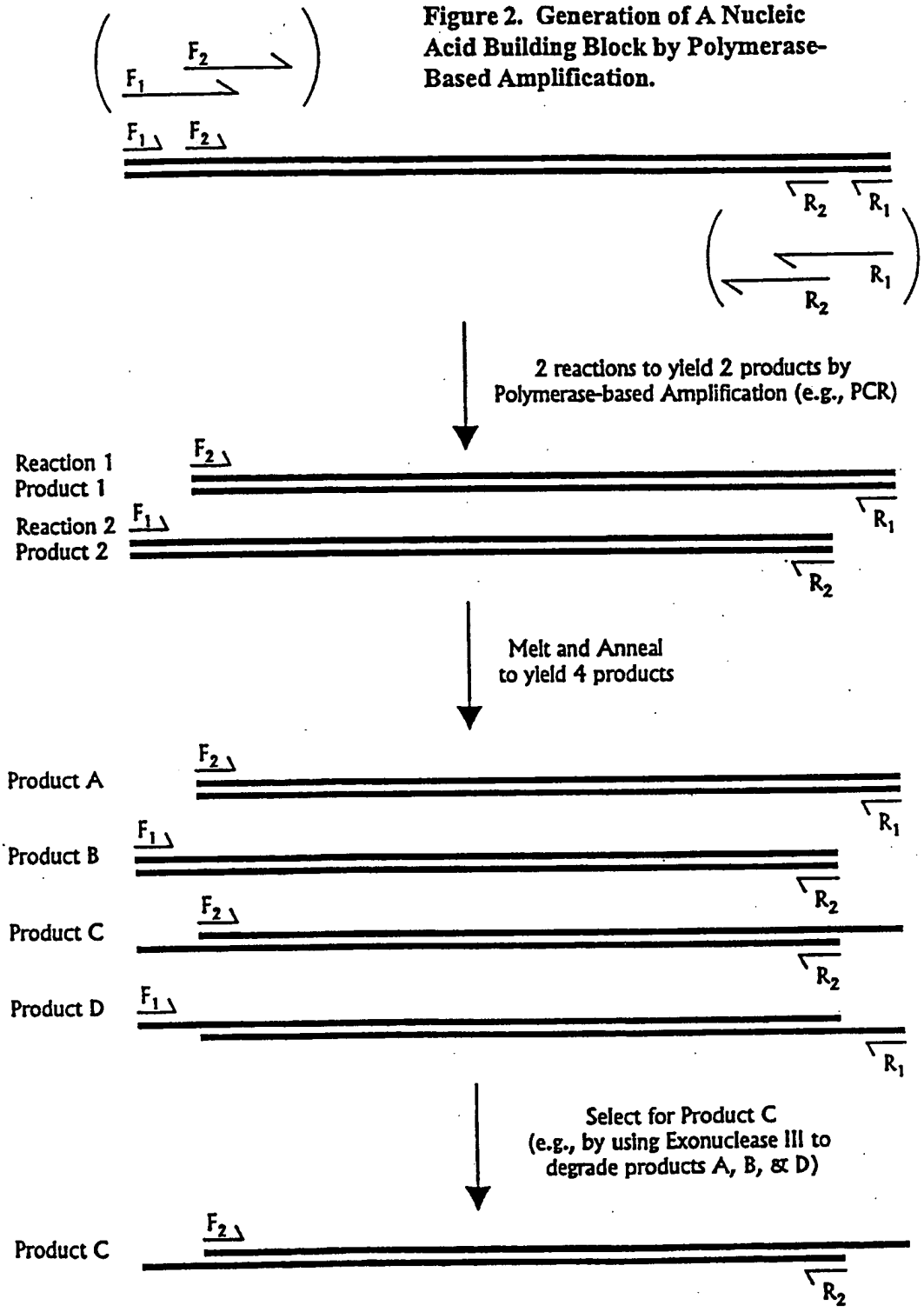
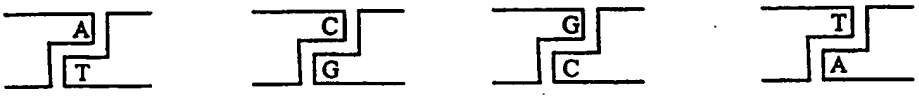


Figure 2. Generation of A Nucleic Acid Building Block by Polymerase-Based Amplification.

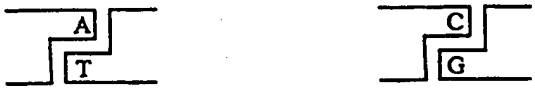
FIGURE 3. Unique Overhangs And Unique Couplings.

The number of unique overhangs of each size (e.g. the total number of unique overhangs composed of 1 or 2 or 3, etc. nucleotides) exceeds the number of unique couplings that can result from the use of all the unique overhangs of that size. For example, the total number of unique couplings that can be made using all the 8 unique single-nucleotide 3' overhangs and single-nucleotide 5' overhangs is 4.

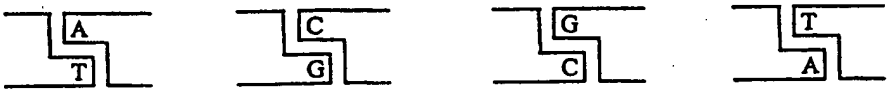
PANEL A. 4 unique single-nucleotide 3' overhangs are possible (i.e., A, C, G, & T). For each of these there is a complementary 3' overhang with which it can pair (i.e., T, G, C, & A, respectively), as shown.



PANEL B. However, the number of unique single-nucleotide 3' overhangs is greater than the number of unique couplings. Thus, only 2 intrintsically unique couplings exist using single-nucleotide 3' overhangs as shown.



PANEL C. Likewise, 4 unique-single nucleotide 5' overhangs are possible (i.e., A, C, G, & T). For each of these there is a complementary 5' overhang with which it can pair (i.e., T, G, C, & A, respectively), as shown.

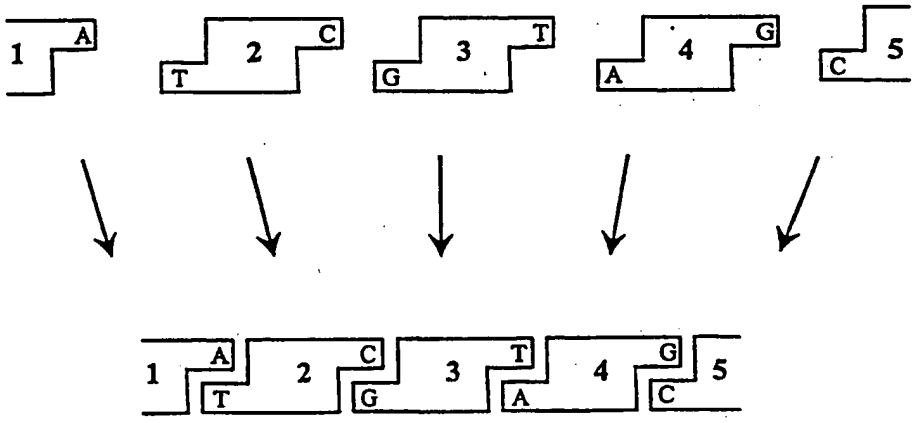


PANEL D. However, the number of unique single-nucleotide 5' overhangs is greater than the number of unique couplings. Thus, only 2 intrinsically unique couplings exist using single-nucleotide 5' overhangs as shown.

FIGURE 4. Unique Overall Assembly Order Achieved by Sequentially Coupling the Building Blocks

Awareness of the degeneracy (between the number of unique overhangs and the number of unique couplings) is important in order to avoid the production of degeneracy in the overall assembly order of the finalized nucleic acid. However, a unique overall assembly order can also be achieved - despite the use of non-unique couplings - by using building blocks having distinct combinations of couplings, and/or by stepping the assembly of the building blocks in a deliberately chosen sequence.

PANEL A. For example, one could attempt to assemble the following nucleic acid product using the 5 nucleic acid building blocks as shown.



PANEL B. However, degeneracy in the overall assembly order of the 5 nucleic acid building blocks would be present if the assembly process were carried out in one step. For example, building block #2 and building block #3 could both couple to building block #1 as shown.

FIGURE 4 cont.

PANEL C. However, a unique overall assembly order could be achieved by sequentially coupling the building blocks in 2 steps (rather than all at once) as shown.

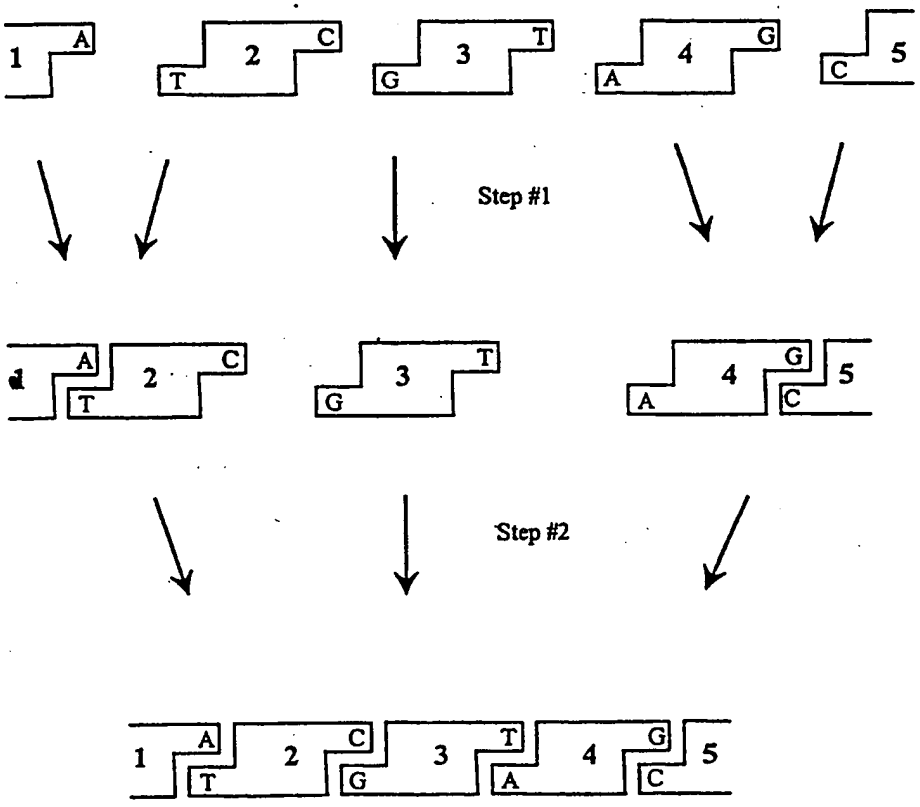
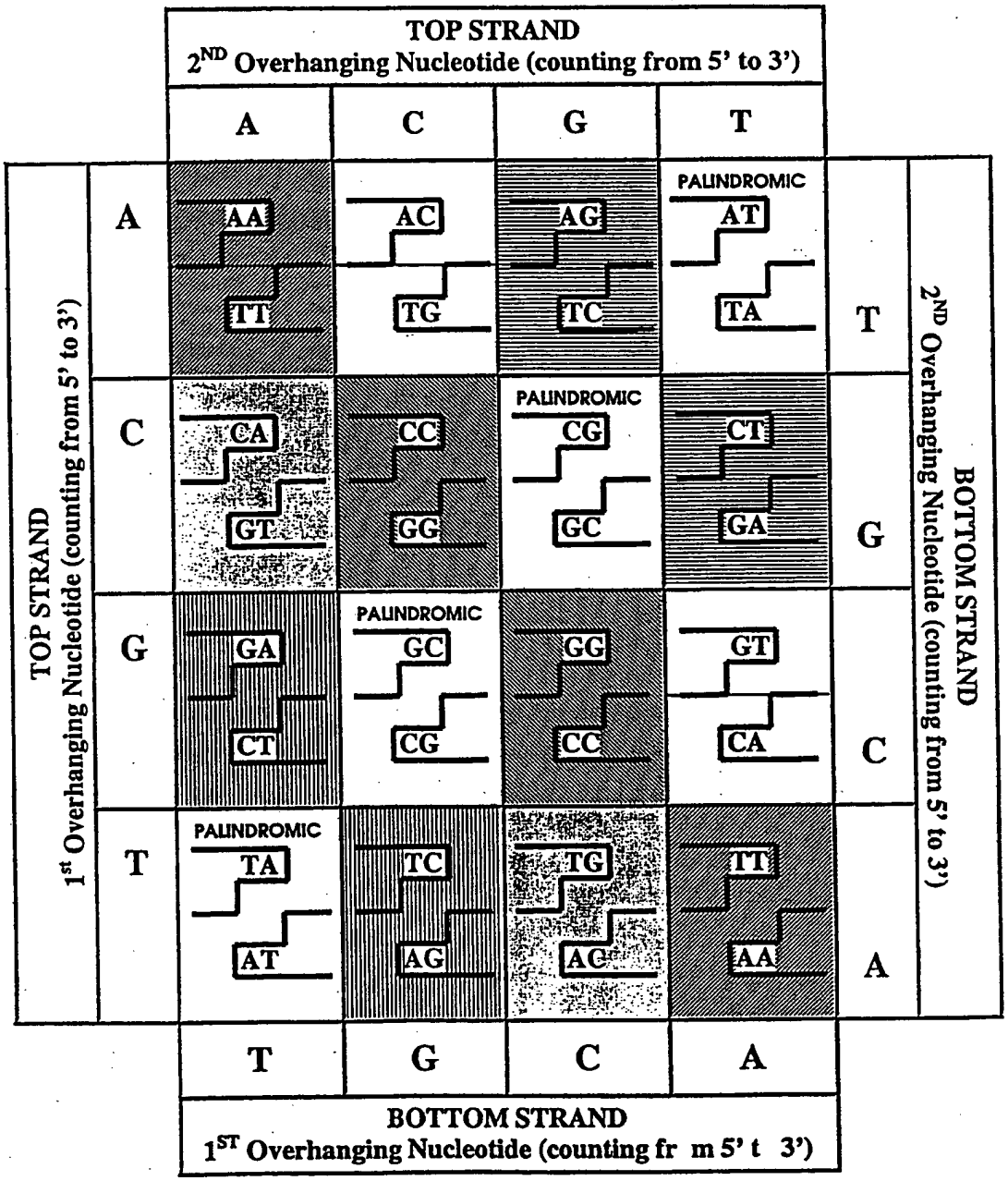


Figure 5. Unique Couplings Available Using a Two-Nucleotide 3' Overhang.

16 unique 3' overhangs can be formed using two-nucleotides. However, use of these 16 unique overhangs allows for the formation of only 6 unique couplings. Another 6 unique couplings are provided by the use 5' overhangs formed using two-nucleotides. Thus, a total of 12 unique couplings are provided by the combined use of 3' and 5' two-nucleotide overhangs. "Twin" couplings are marked in the same shading.



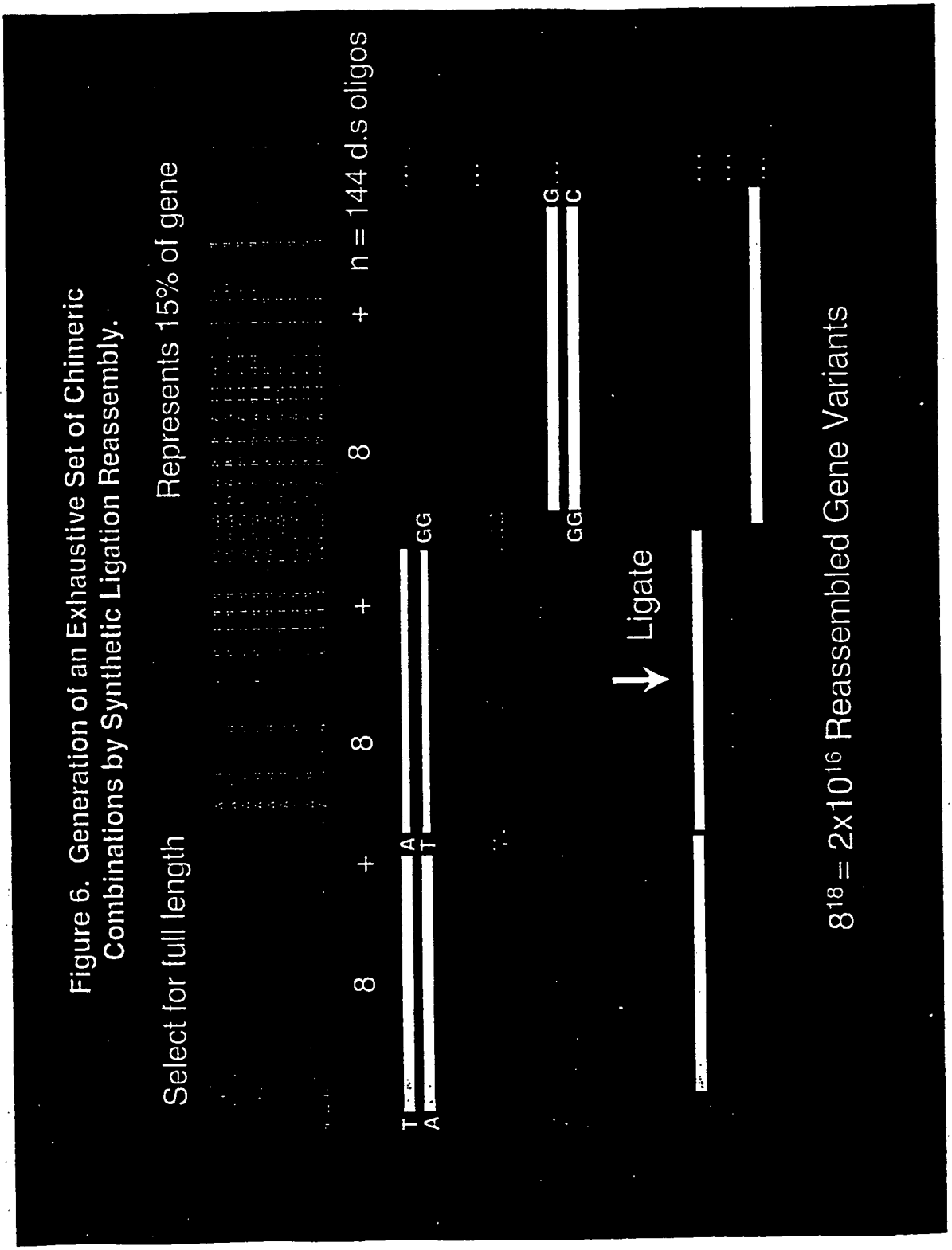


Figure 6. Generation of an Exhaustive Set of Chimeric Combinations by Synthetic Ligation Reassembly.
Represents 15% of gene
Select for full length
8 + 8 + 8 + n = 144 d.s oligos
T
A
A
T
GG
GG
G...
C
Ligate
$8^{18} = 2x10^{16}$ Reassembled Gene Variants

Figure 7. Synthetic genes from oligos.

```
                 NcoI                                                                CCGT
150am13_00       c ATGATGCACG GCGATATTTC ATCGAGCAAT GACACGGTCG GCGTTGCCGT
150AM7_001       c ATGCATCACG GCGACATTTC ATCGAGCAAT GACACGGTCG GCGTTGCCGT
431am7_002       c ATGAGACACG GAGATATCTC CAGCAGCAAC GATTGCGTGG GCGTGGCCGT

                                                                GAG  GT
150am13_00       CGTGAACTAC AAGATGCCTC GCCTTCATAC CAAGGCGGAG GTTTTAGCGA
150AM7_001       CGTGAACTAC AAGATGCCGC GGCTTCACAC CAAGGCTGAG GTGCTGGCCA
431am7_002       CGTGAACTAC AAGATGCCGC GGCTGCATAC CCGCGCGGAG GTGATGGAGA

                                                           CGG
150am13_00       ACGCCAGAAA GATCGGCGAG ATGATCGTCG GCATGAAGAC CGGCCTGCCC
150AM7_001       ACTGCCGCAA GATCGCCGAC ATGCTGGTCG GCATGAAGAG CGGCCTGCCG
431am7_002       ACGCCCGCAA GATCGCCGAC ATGGTCGTGG GCATGAAGCG CGGCCTGCCC

                                                      CCACG
150am13_00       GGAATGGATC TGGTGATCTT CCCGGAATAT TCGACCCACG GCATCATGTA
150AM7_001       GGAATGGATC TGGTGATCTT CCCGGAATAT TCCACCCACG GCATCATGTA
431am7_002       GGCATGGACC TGGTCATCTT CCCCGAGTAC TCCACCCACG GCATCATGTA

                                                                CCC  GG
150am13_00       CGACTCCAAG GAAATGTACG ATACCGCGTC CGTCGTGCCC GGCGAGGAGA
150AM7_001       CGACTCCAAG GAGATGTACG ACACGGCGTC GACGGTGCCG GGTGAAGAGA
431am7_002       CGACGCCAAG GAAATGTACG AAACCGCTTC GGCCATTCCG GGCGAAGAGA

                                                                  G  GGG
150am13_00       CCGAGATTTT TGCCGAAGCC TGCCGCAAGG CGAAAGTCTG GGGCGTGTTC
150AM7_001       CCGAGATTTT CGCCGAGGCC TGCCGCAAGG CCAAGGTCTG GGGCGTGTTC
431am7_002       CTGCTGTGTT CGCCGACGCC TGCCGCAAGG CCAACGTATG GGGCGTGTTT

                                                            AAAG  C
150am13_00       TCGCTCACCG GCGAACGTCA CGAGGAACAT CCGAAGAAGG CGCCCTACAA
150AM7_001       TCGCTGACCG GCGAGCGCCA CGAGGAGCAT CCCAATAAAG CGCCGTACAA
431am7_002       TCGCTGACGG GCGAGCGCCA CGAAGAGCAC CCGAACAAGG CGCCGTACAA

                                                                CAG  AA
150am13_00       CACGCTGATC CTGATGAACG ACAAGGGCGA GGTGGTCCAG AAATACCGCA
150AM7_001       CACCCTGATC CTGATGAACG ACAAGGGTGA AGTCGTTCAG AAATATCGCA
431am7_002       CACGCTCATC CTGATGAACA ACAAGGGCGA GATCGTCCAG AAGTACCGCA

                                                       GGTA
150am13_00       AGATCATGCC GTGGGTTCCG ATCGAGGGCT GGTACCCCGG CAACTGCACC
150AM7_001       AGATCATGCC GTGGGTGCCG ATCGAAGGCT GGTATCCCGG CAACTGCACG
431am7_002       AGATCATGCC CTGGGTGCCG ATCGAAGGCT GGTATCCGGG CGATTGCACC

                                                  TGAAG
150am13_00       TACGTCTCCG ACGGGCCGAA GGGCATGAAG GTTTCGCTGA TCATCTGCGA
150AM7_001       TACGTCTCCG AAGGCCCGAA GGGCATGAAG ATGTCGCTGA TCATCTGCGA
431am7_002       TATGTGTCGG AAGGCCCCAA GGGACTGAAG ATCAGCCTCA TCATCTGCGA

                                                  TCTGGCG
150am13_00       TGACGGCAAC TATCCGGAAA TCTGGCGCGA CTGCGCCATG AAGGGCGCCG
150AM7_001       CGACGGCAAC TACCCGGAAA TCTGGCGTGA CTGCGCGATG AAGGGCGCCG
431am7_002       CGACGGCAAT TACCCCGAGA TCTGGCGCGA TTGCGCCATG CGCGGCGCCG
```

Figure 7 cont.

```
                                  CCAG
150am13_00      AGCTGATCGT GCGCTGCCAG GGCTACATGT ATCCGGCCAA GGACCAGCAG
150AM7_001      AACTGATCAT CCGCTGCCAG GGCTACATGT ATCCCGCCAA GGATCAGCAG
431am7_002      AGCTGATCGT GCGTTGCCAG GGATACATGT ACCCGGCCAA GGACCAGCAG

                                  GC
150am13_00      GTCATCATGG CGAAGGCGAT GGCGTGGGCG AATAATTGTT ACGTCGCGGT
150AM7_001      GTGCTGATGG CGAAAGCAAT GGCCTGGGCC AACAACGTTT ATGTCGCGGT
431am7_002      GTCATGGTGT CCAAGGCCAT GGCGTGGATG AACAACGTCT ACGTGGCGGT

                              GGGCTTCG
150am13_00      TTCCAATGCC GCGGGCTTCG ATGGCGTCTA TTCGTATTTC GGCCACTCGG
150AM7_001      CGCCAATGCC TCGGGCTTCG ACGGCGTCTA CTCGTATTTC GGCCATTCGG
431am7_002      GGCCAATGCC GCGGGCTTCG ACGGCGTGTA TTCCTACTTC GGCCATTCGG

                             TTCGA
150am13_00      CGATCATCGG CTTCGATGGC CGCACGCTCG GCGAATGCGG CGAGGAAGAA
150AM7_001      CGATCATCGG CTTCGACGGC CGTACCCTCG GCGAATGCGG CGAGGAGGAT
431am7_002      CCATCATCGG CTTCGACGGC CGCACGCTGG GCGAATGCGG TGAAGAAGAC

                         C AGTA
150am13_00      TACGGCATCC AGTATGCCCA GCTTTCGAAG ATGCTGATCC GCGACGCCCG
150AM7_001      TATGGCATCC AGTATGCCGC CATCTCCAAG TCGCTGATCC GCGACGCGCG
431am7_002      ATGGGCGTGC AGTACGCCGA GCTCTCCACC AGCCTGATCC GCGACGCGCG

                             CAATC
150am13_00      CCGCACCGGA CAATCGGAAA ACCATCTCTT CAAGCTGGTG CATCGTGGCT
150AM7_001      CCGCACCGGC CAATCGGAAA ACCATCTCTT CAAGCTGGTG CACCGTGGCT
431am7_002      CAAGAACATG CAGTCGCAGA ACCACTTGTT CAAGCTGGTG CACCGCGGCT

                             GATCAA
150am13_00      ACACCGGGTT GATCAACTCC GGCGAGGGCG ACCGCGGTCT CGCGGCCTGT
150AM7_001      ACACCGGCAT GATCAATTCC GGCGAGGGCG ACCGCGGTGT CGCGGCTTGC
431am7_002      ACACCGGCAA GATCAATTCC GGCGAAGAGG CCACCGGCGT CGCGGCATGC

                        TTA
150am13_00      CCTTATGAGT TCTACAACAA ATGGATCGCC GATCCGGAAG GCACCCGCGA
150AM7_001      CCGTATGATT TCTATTCGAA ATGGATCGCC GATCCCGAGG GTACACGCGA
431am7_002      CCGTACAACT TCTACGCCAA CTGGATCAAC GATCCGGAGG GCACGCGCAA

                       ATGGT
150am13_00      AATGGTCGAG TCCTTTACCC GGCCGACGGT GGGAACCGAT GAAGCGCCCA
150AM7_001      GATGGTGGAA TCCTTCACGC GTCCGACGGT GGGTGTGGAG GAATGCCCGA
431am7_002      GATGGTCGAA TCCTTCACCC GGTCCACCGT GGGCACGCCG GAGTGCCCCA

                      TCGAG
150am13_00      TCGAAGGCAT CCCGAACAAG GTCGCGGTGC ACCGCTGA   aagct
150AM7_001      TCGAGGGCAT TCCGAACAAG GCCACCACGC ACCGCTGA   aagct
431am7_002      TGGACGGCAT CCCCAACGAG GACGCCAAGC ACCGCTAG   aagct
                                                            HindIII
```

Figure 8. Nucleic acid building blocks for synthetic ligation gene reassembly.

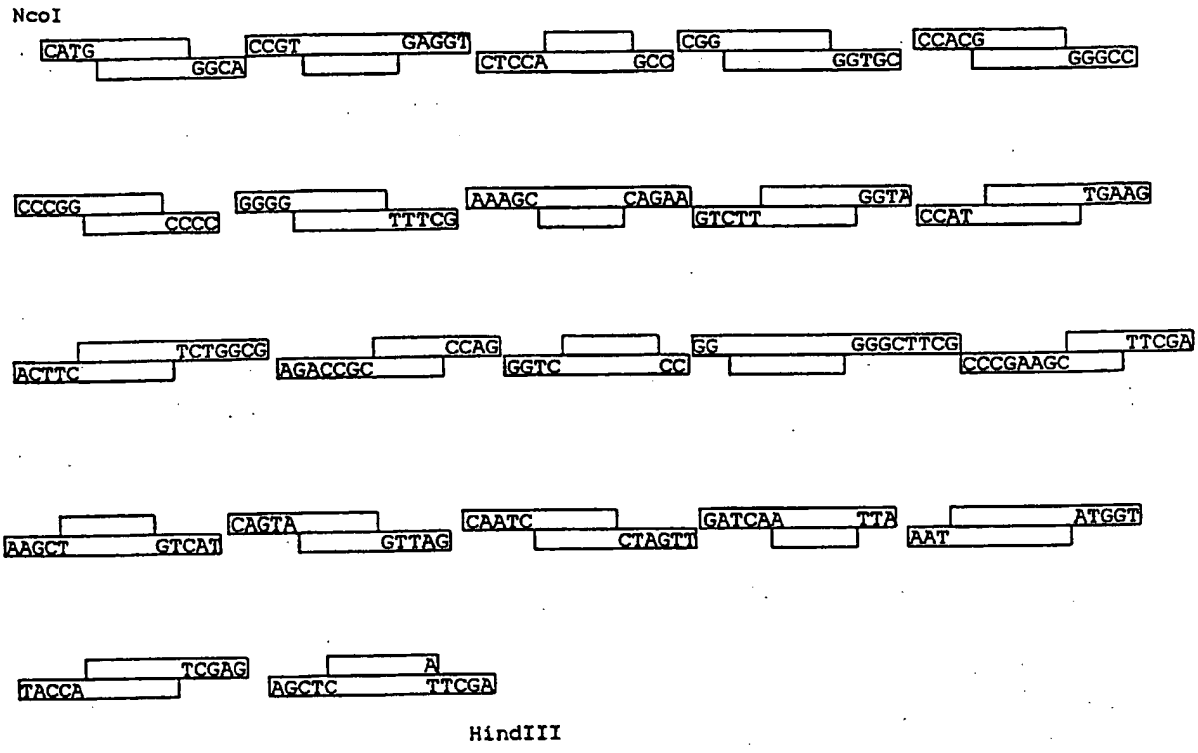
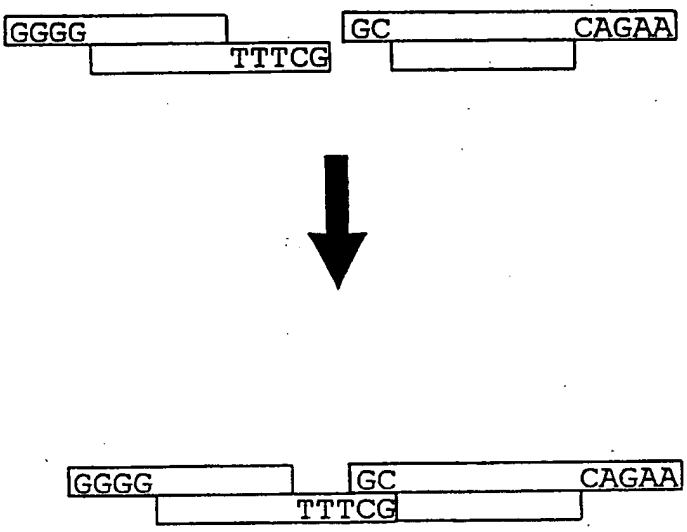


Figure 9. Addition of Introns by Synth tic Ligation Reassembly.

Figure 10. Ligation Reassembly Using Fewer Than All The Nucleotides Of An Overhang.

Gap Ligation



Ligation of one strand only;
gap in second strand can be repaired in vivo

Figure 11. Avoidance of unwanted self-ligation in palindromic couplings.

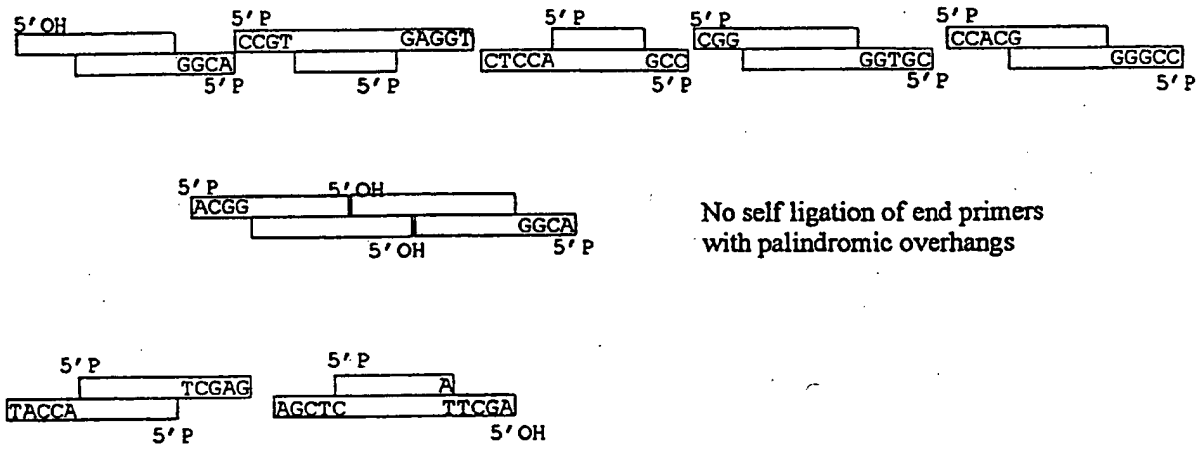
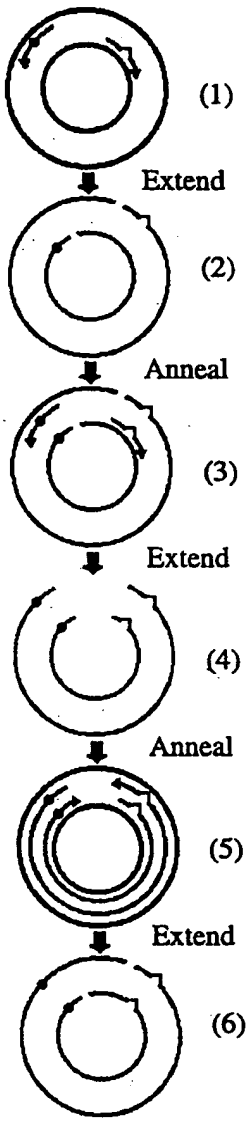


Figure 12

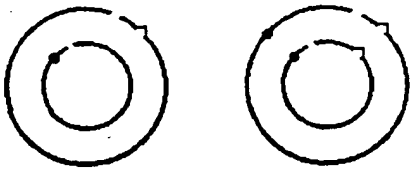
Site-Directed Mutagenesis by Polymerase-based Extension

Panel A



Panel B

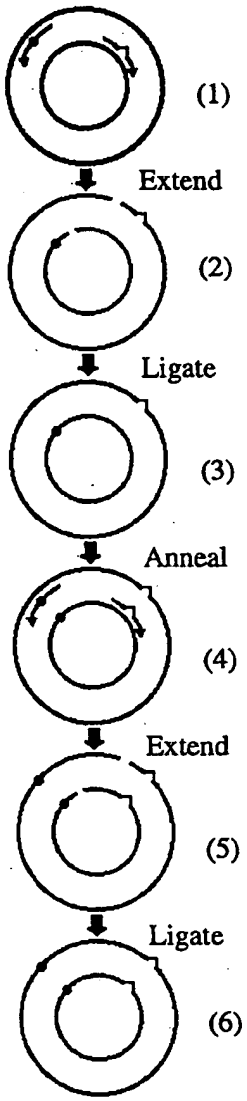
Amplification products are comprised of the following molecular structures:



Molecule (A) Molecule (B)

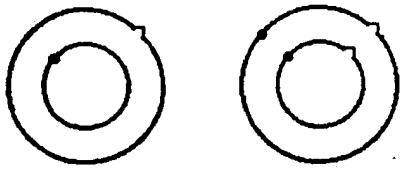
Figure 13
Site-Directed Mutagenesis By Polymerase-based Extension and Ligase-based Ligation

Panel A



Panel B

Amplification products are comprised of the following molecular structures:



Molecule (A) Molecule (B)

Figure 14

Strategy for obtaining and using nucleic acid binding proteins that facilitate entry of genetic vaccines.

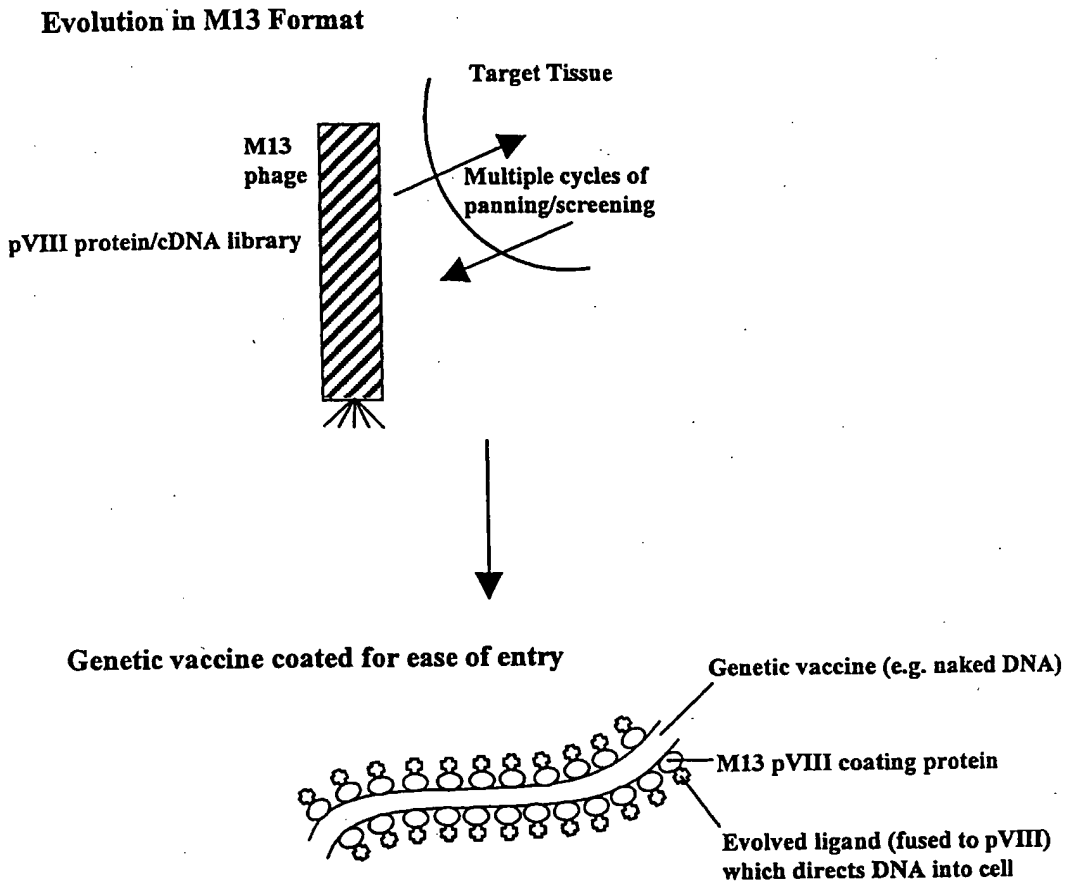


Figure 15

A schematic representation of a method for evolving a chimeric, multivalent antigen that has immunogenic regions from multiple antigens.

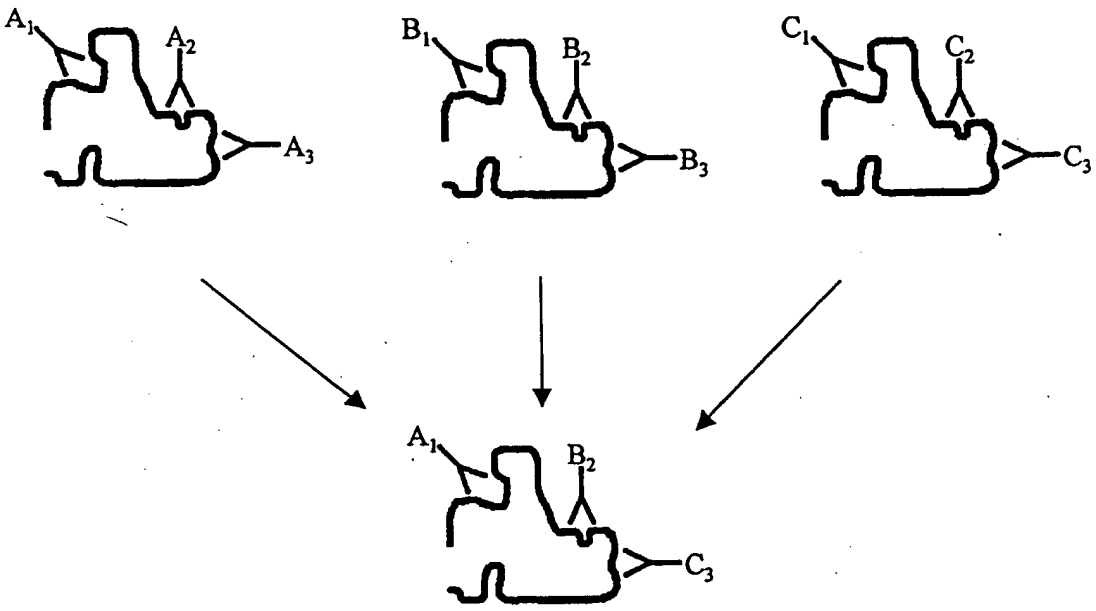


Figure 16

Method for Obtaining Non-Stochastically Generated Polypeptides that can induce a Broad-Spectrum Immune Response.

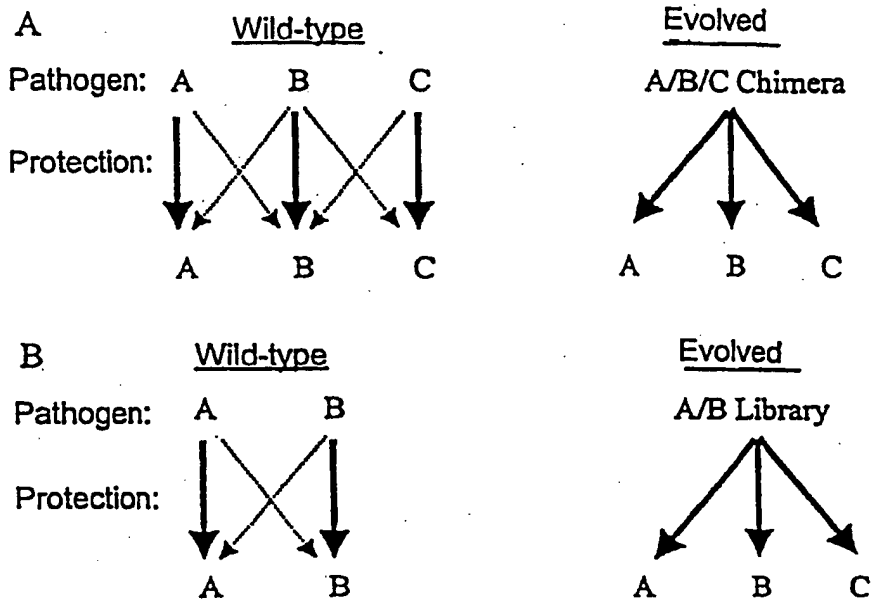


Figure 17

Possible factors for determining whether a particular polynucleotide encodes an immunogenic polypeptide having a desired property.

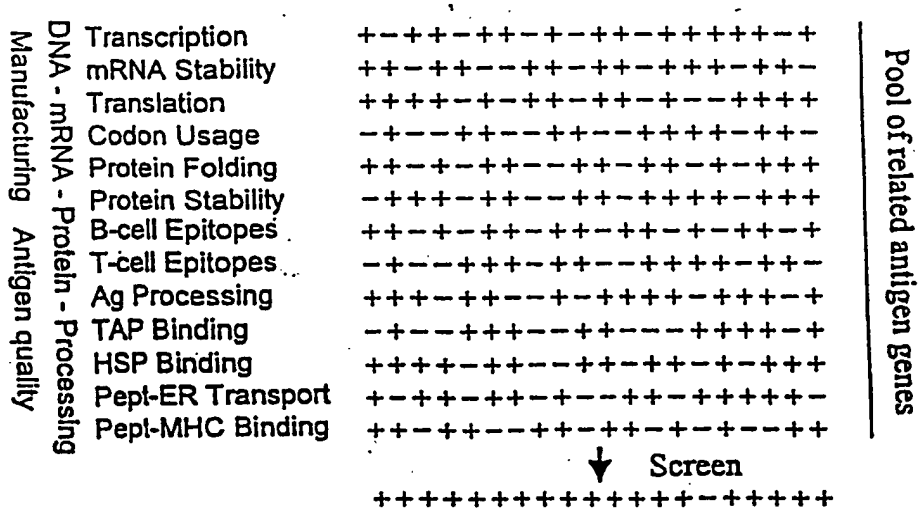


Figure 18

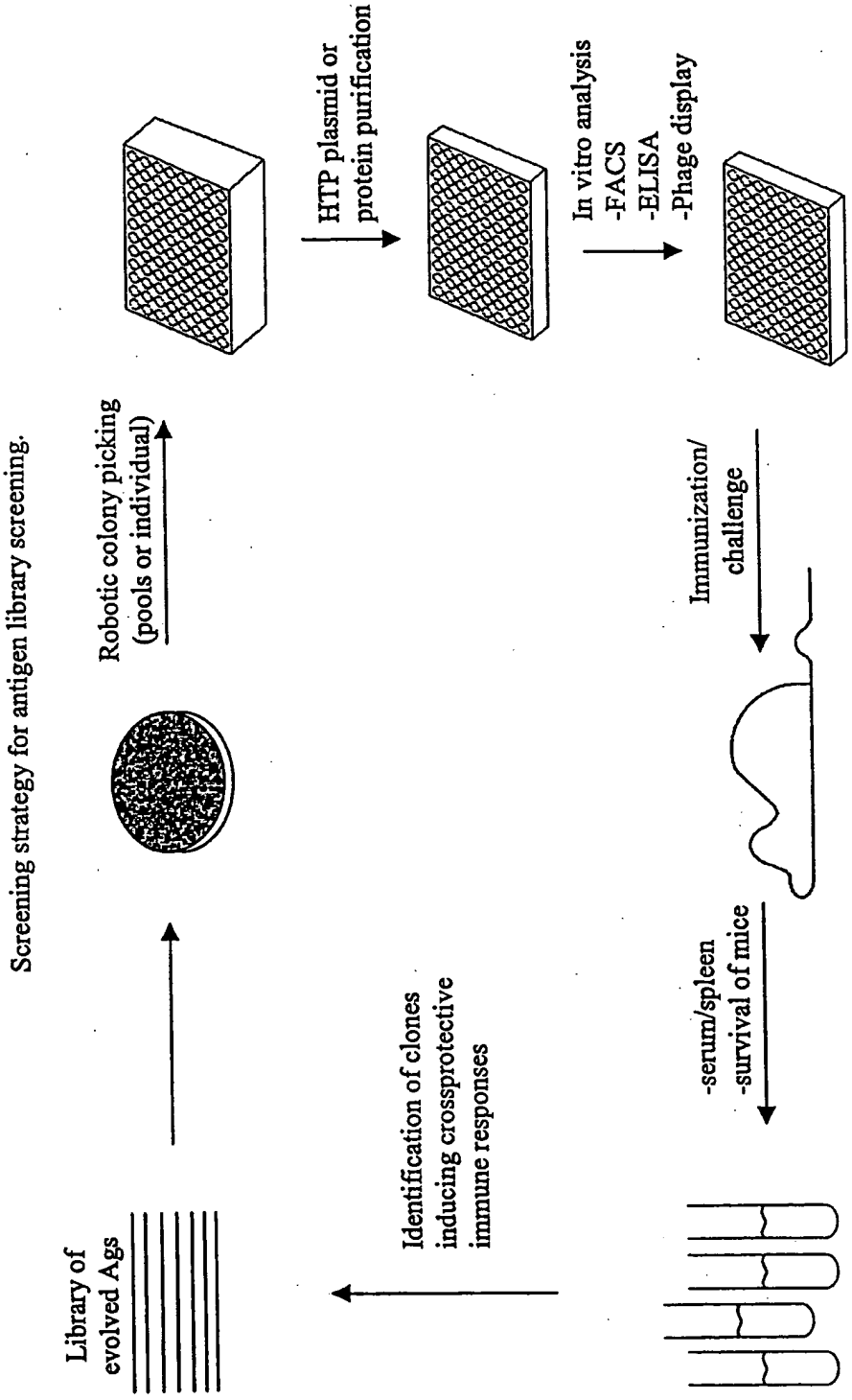

Screening strategy for antigen library screening.
Robotic colony picking (pools or individual)
HTP plasmid or protein purification
In vitro analysis -FACS -ELISA -Phage display
Immunization/ challenge
-serum/spleen -survival of mice
Identification of clones inducing crossprotective immune responses
Library of evolved Ags

Figure 19

Strategy for pooling and deconvolution as used in antigen library screening

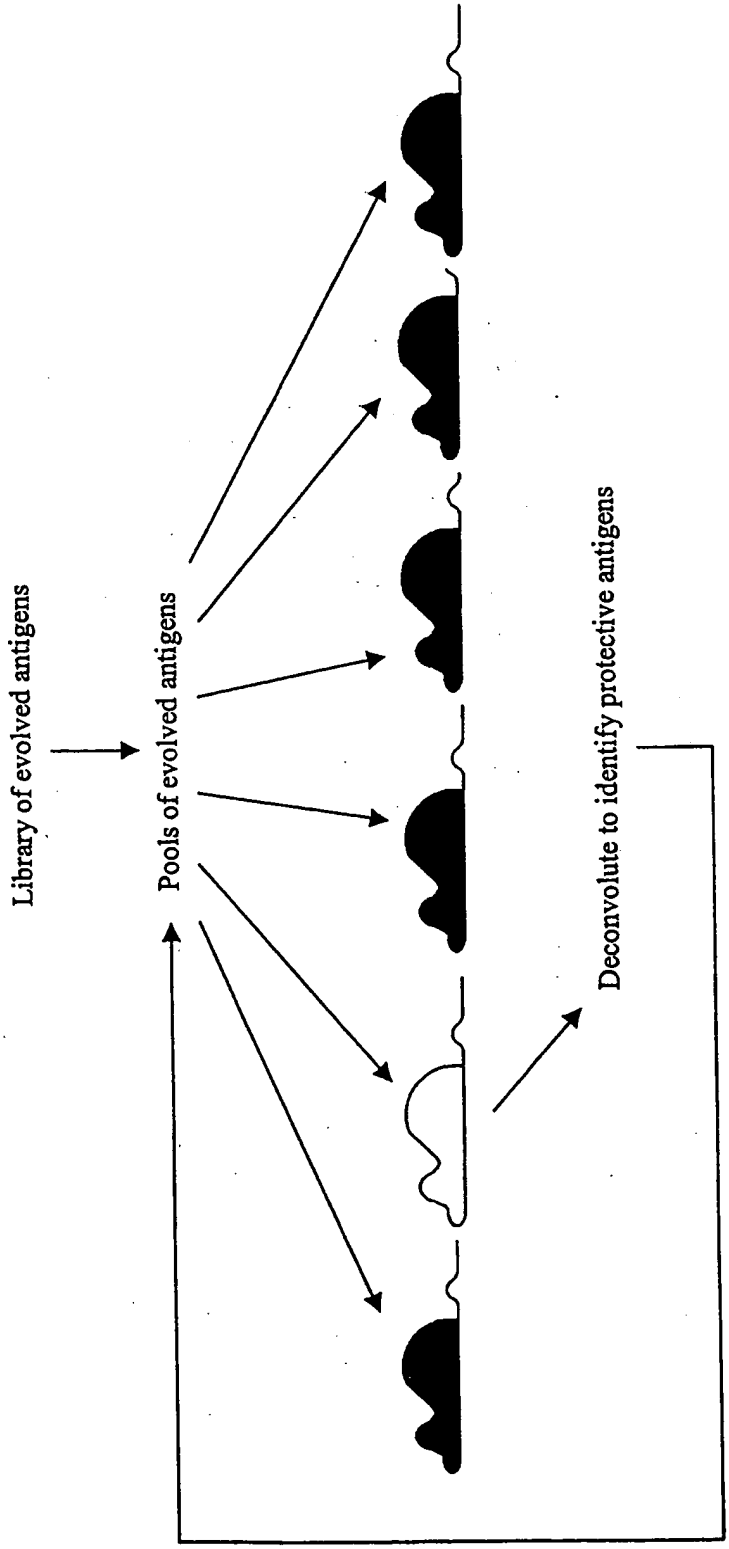
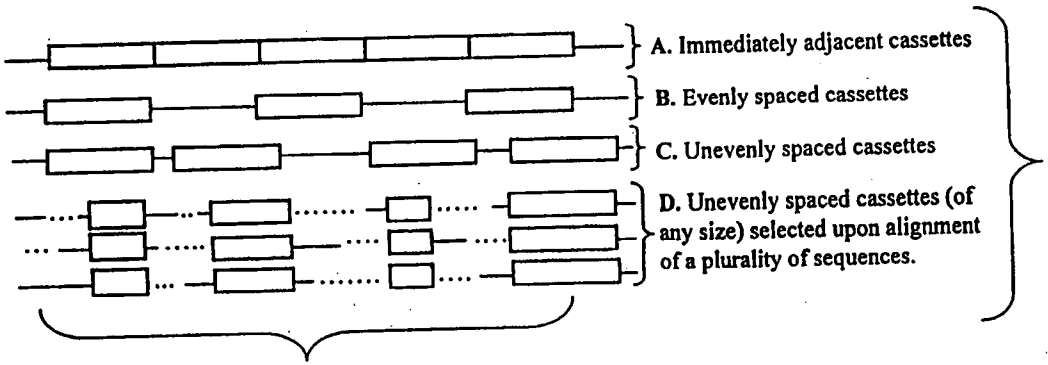


Figure 20. Preferred embodiments of site-saturation mutagenesis



I. SEQUENCES TO BE MUTAGENIZED

Nucleic acid sequences that can be mutagenized by non-stochastic site-saturation mutagenesis include any polynucleotide of 15-100,000 bases in length (e.g. any coding region & any non-coding region).

Specific examples:

- Promoter
- Enhancer
- Intron
- Whole gene
- cDNA
- Gene segment
- Operator
- Polynucleotide functional group
- Expression regulatory sequence
- Expressed sequence tag (EST)
- Open reading frame (ORF)
- Repressor/transactivator
- Origin of replication
- Gene pathway

II. MUTAGENIC CASSETTES WITHIN SEQUENCE TO BE MUTAGENIZED ( )

Mutagenic cassettes that can be mutagenized by non-stochastic site-saturation mutagenesis include any polynucleotide cassette of 1-500 bases in length. Site-saturation mutagenesis is servicable for mutagenizing a complete set of cassettes contained within a polynucleotide sequence to be mutagenized. As shown, cassettes can be spaced along each polynucleotide differently (i.e. immediately adjacent, evenly spaced, or unevenly spaced) and of any size. In a preferred but non-limiting exemplification a set of mutagenic cassettes is a set of contiguous codons within a sequence of defined length. Alternatively, in another preferred but non-limiting example, a set of mutagenic cassettes is a set of nucleotide cassettes that are not shared by aligned related polynucleotides.

III. TYPES OF MUTATIONS THAT CAN BE INTRODUCED INTO MUTAGENIC CASSETTES

The type of mutations to be introduced in a set of mutagenic cassettes can be of the same type or of different types within each round of polynucleotide site-saturation mutagenesis. Each mutagenic cassette (within the nucleic acid sequence to be mutagenized) preferably is usually mutagenized by the use of a corresponding oligo (including by a degenerate oligo). Examples of degenerate mutations provided by this invention include:

- Codons for all 20 amino acids (e.g. N,N,N or N,N,G/T or N,N,G/C)

  All degenerate codons that do not change the amino acid sequence of the parental template (i.e. codons for the same amino acid that is present in the parental template)

  Codons (all or selected) for amino acids within the same grouping according to the selected amino acid grouping scheme*.

- Codons for at least 1 amino acid in each amino acids group*.

*Exemplary amino acid grouping schemes (notes, some groups overlap each other):

- Aromatic (Phe, Trp, Tyr)
- Aliphatic (Gly, Ala, Val, Leu, Ile)
- OH-containing (Ser, Tyr, Thr)
- Acidic (Asp, Glu, Asn, Gln)
- Basic (Lys, Arg, His)
- Sulfur-containing (Met, Cys)
- Polar (Ser, Thr, Cys, Asn, Gln, Tyr)
- Non-polar (Gly, Ala, Val, Leu, Ile, Met, Phe,Trp, Pro)

Figure 21

Schematic representation of a multimodule genetic vaccine vector (relative sizes of functional units are not drawn to scale)

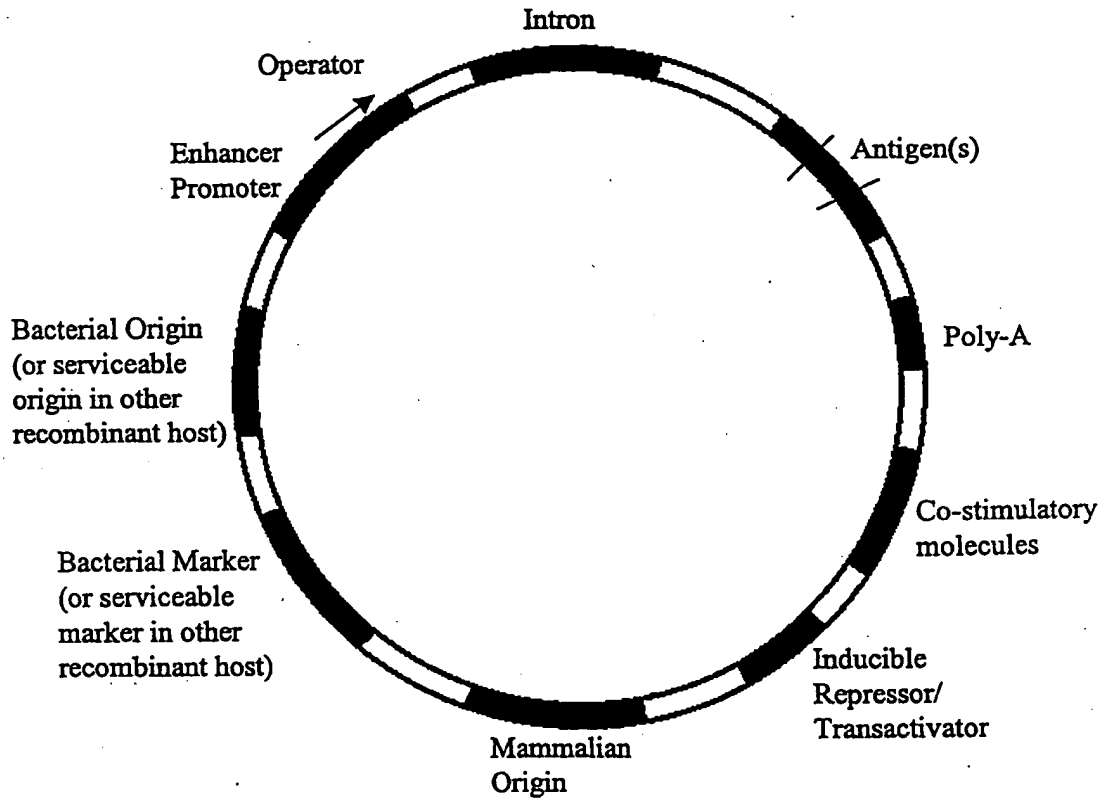


Figure 22A and 22B

Generation of vectors with multiple T cell epitopes.

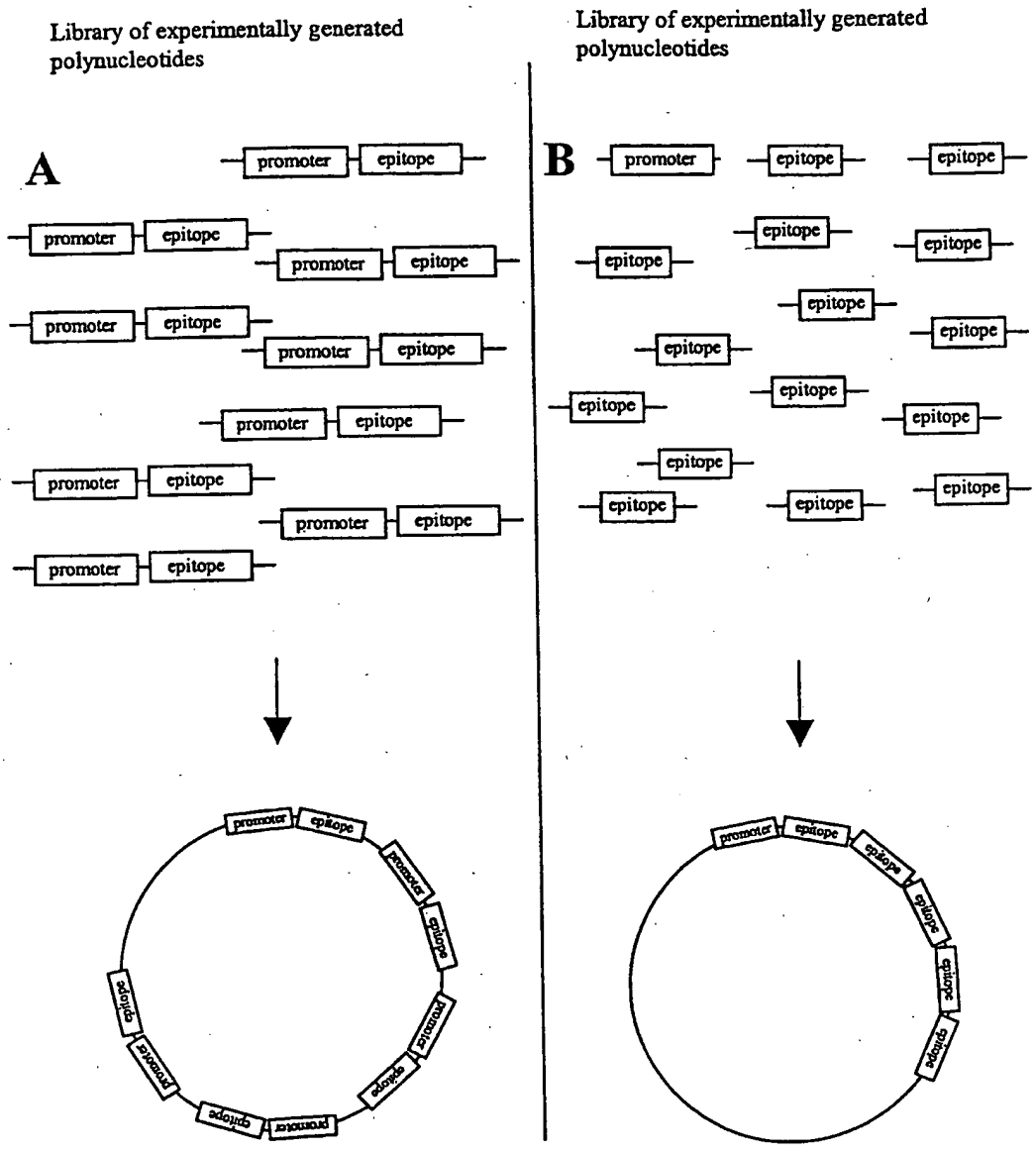


Figure 23

Generation of optimized genetic vaccines by directed evolution

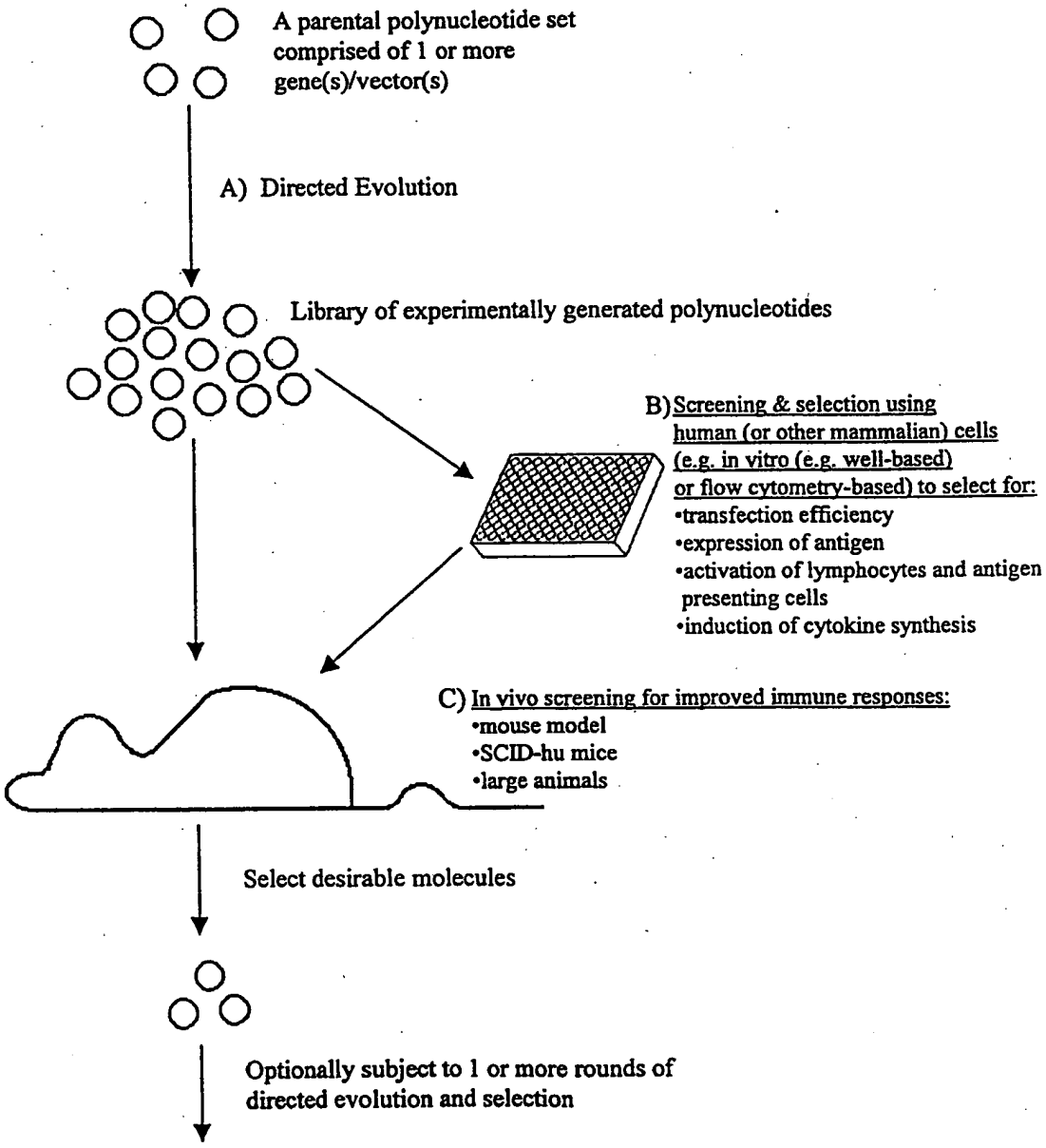


Figure 24

Recursive application of directed evolution and selection of evolved promoter sequences as an example of flow cytometry-based screening methods.

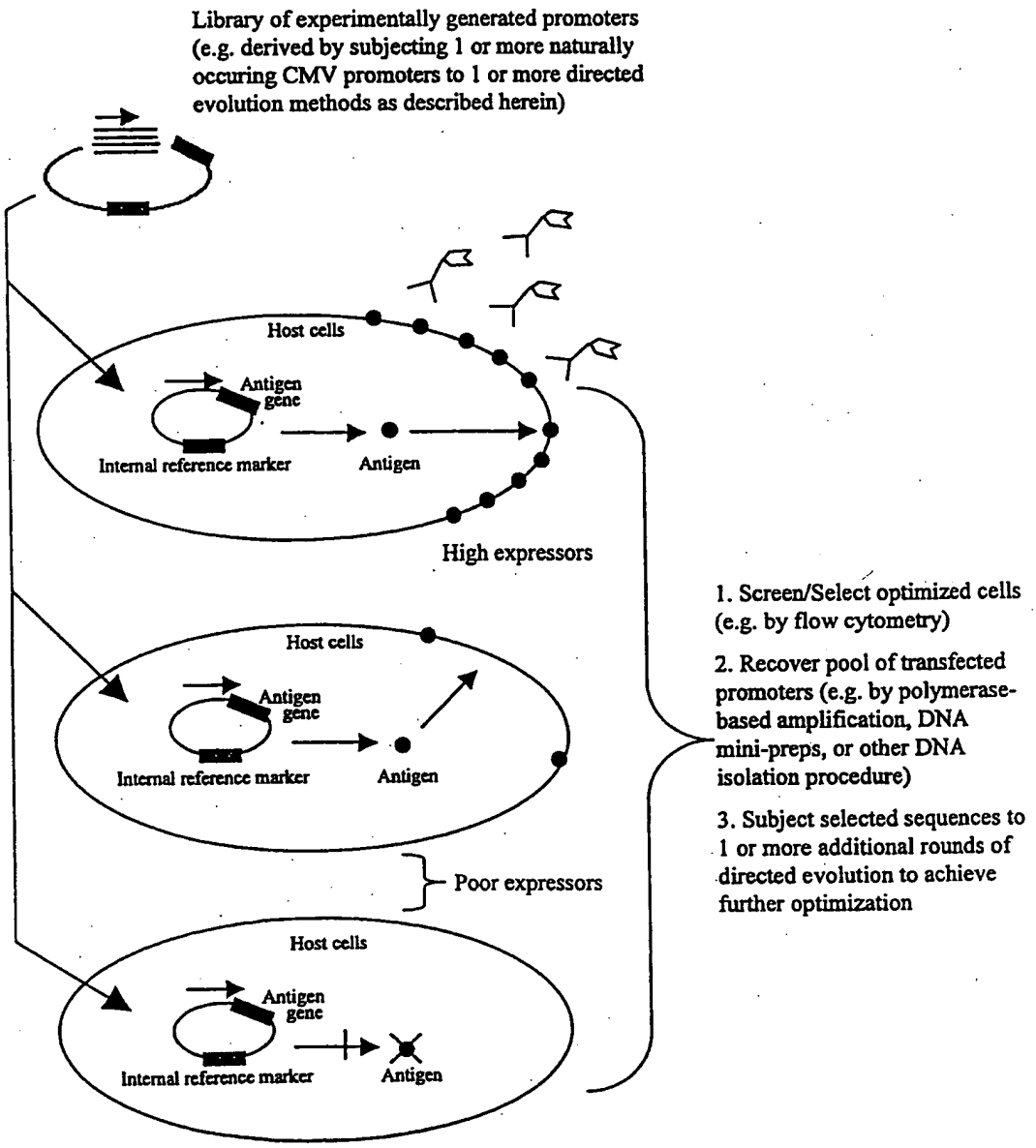


Figure 25

An apparatus for microinjections of skin and muscle.

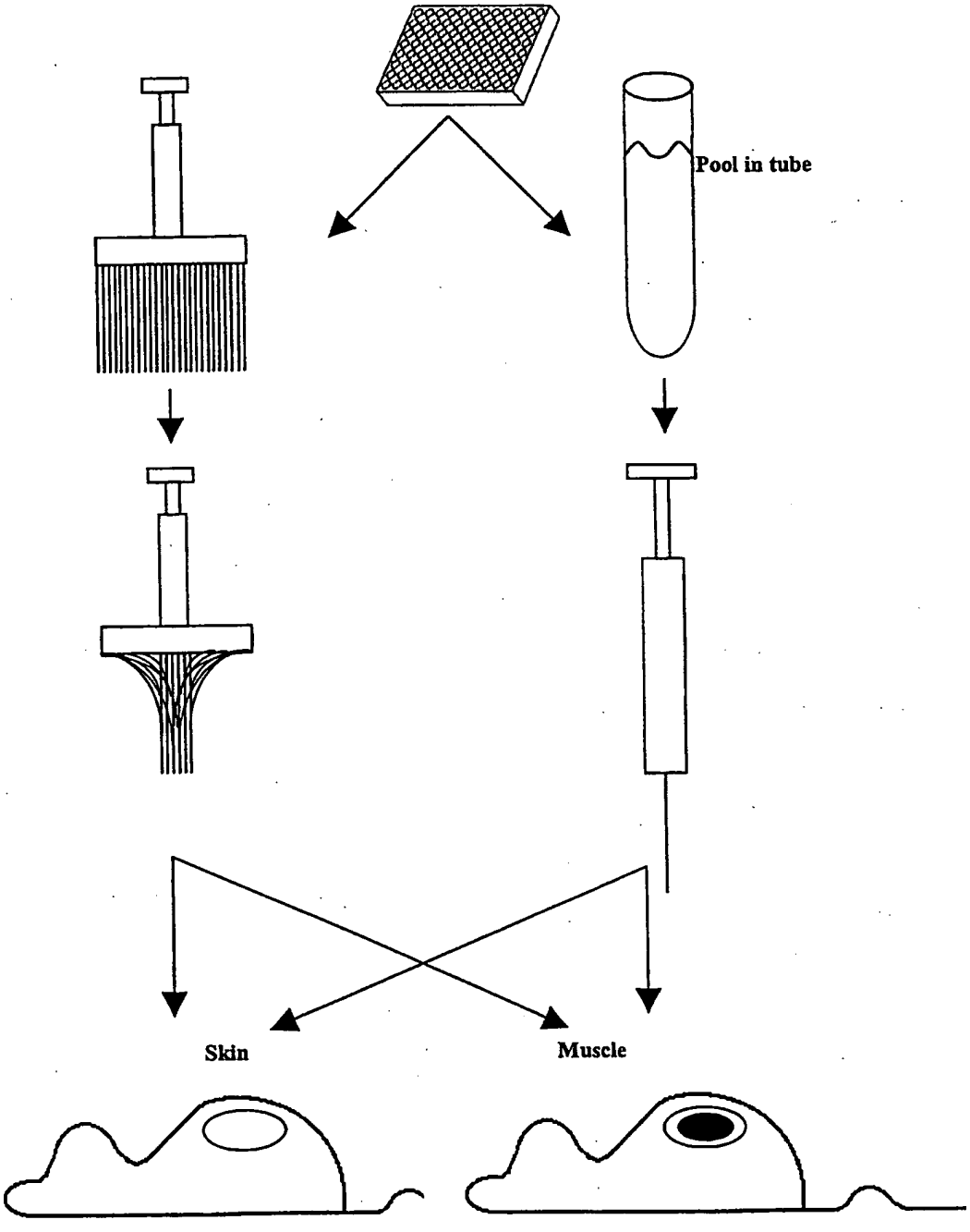


Figure 26 Panel A

Non-stochastic polynucleotide reassembly in combination with non-stochastic polynucleotide site-saturation mutagenesis.

Shown below is a non-limiting example of a permutation of the directed evolution methods described herein

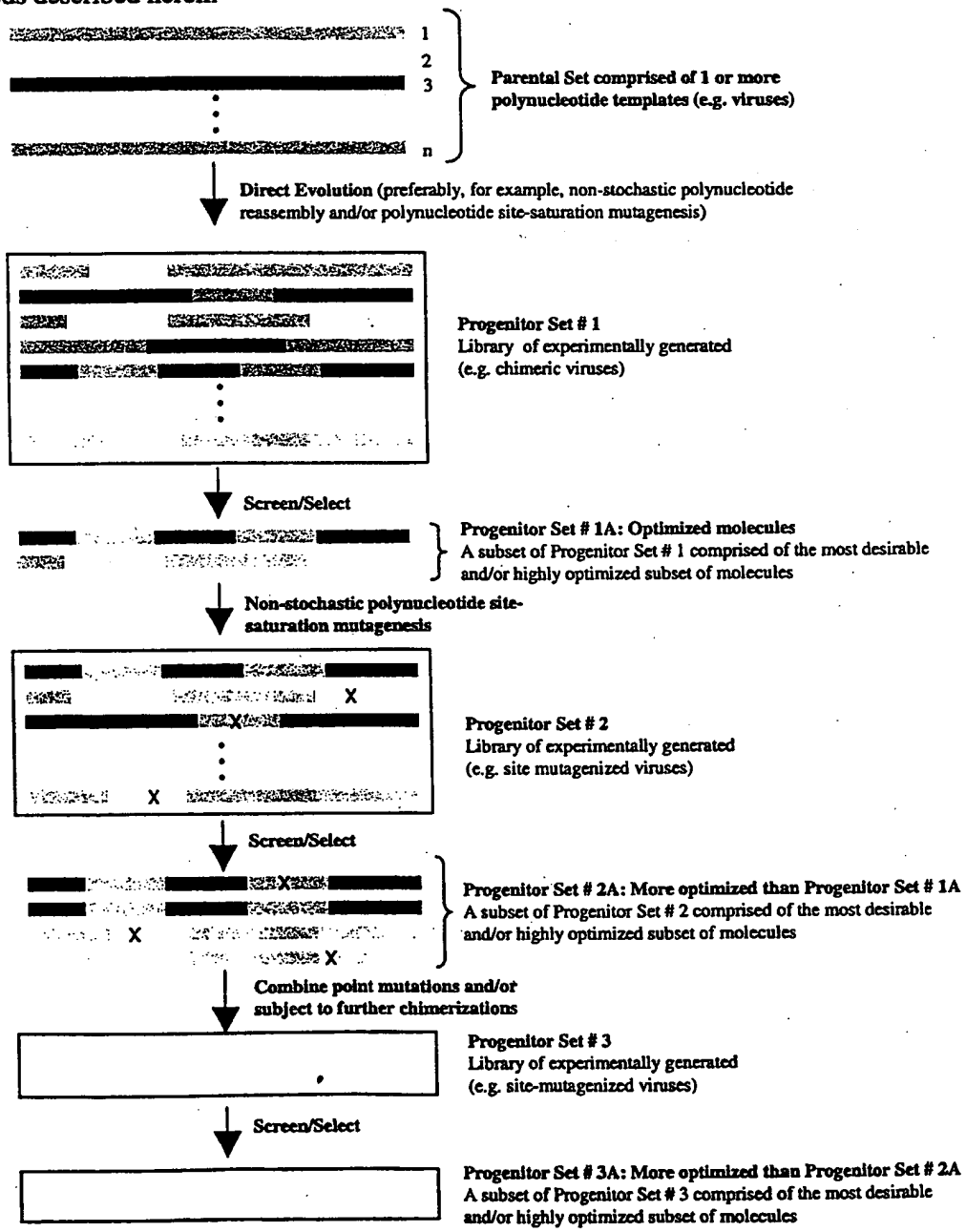


Figure 26 (continued) Panel B

Screening of experimentally generated molecules produced by non-stochastic polynucleotide reassembly in combination with non-stochastic polynucleotide site-saturation mutagenesis

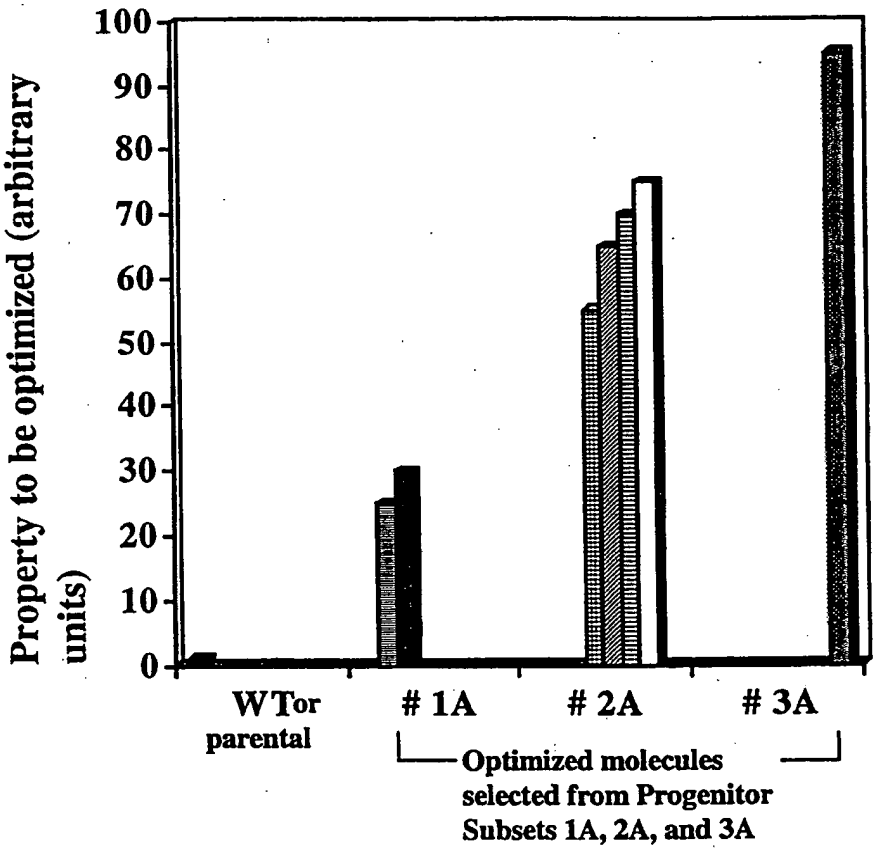


Figure 27

Vector for promoter evolution

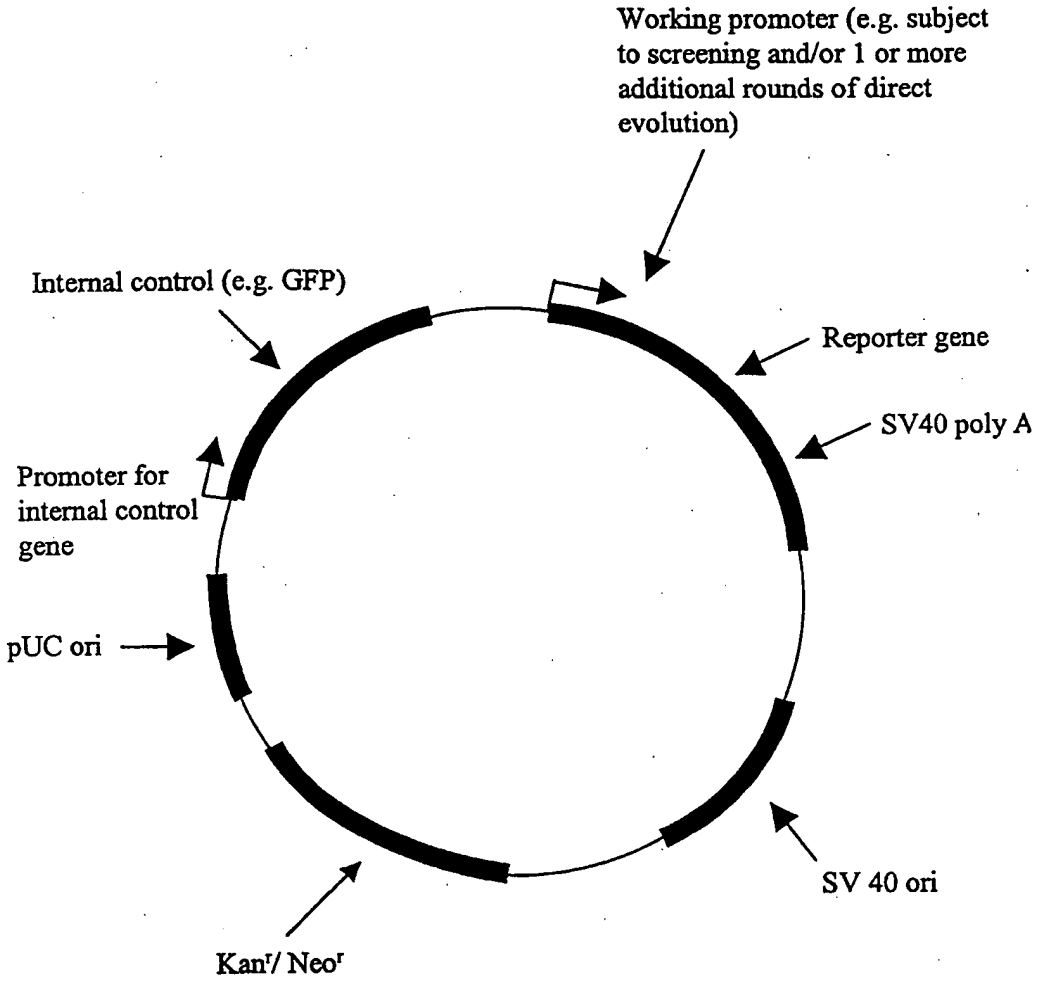


Figure 28

Iterative evolution of inducible promoters using directed evolution and flow cytometry-based selection.

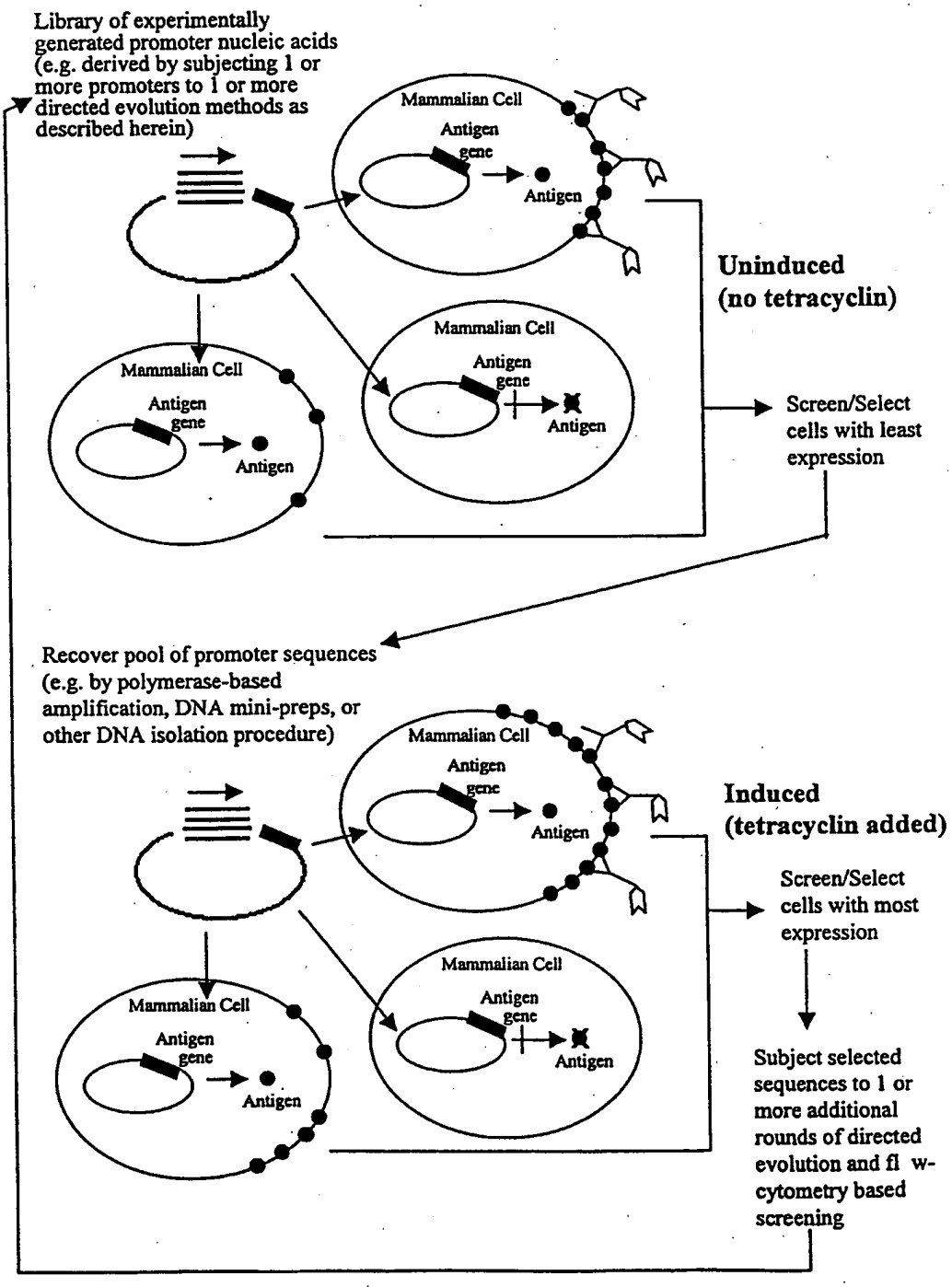


Figure 29

The present invention provides that a genetic vaccine can be subjected to directed evolution in order to achieve improved effectiveness upon administration by oral, intravenous, intramuscular, intradermal, anal, vaginal, or topical delivery methods.

The figure below shows an example of the directed evolution of a genetic vaccine, comprised of an M13 phage-based vaccine, to achieve optimization for oral delivery.

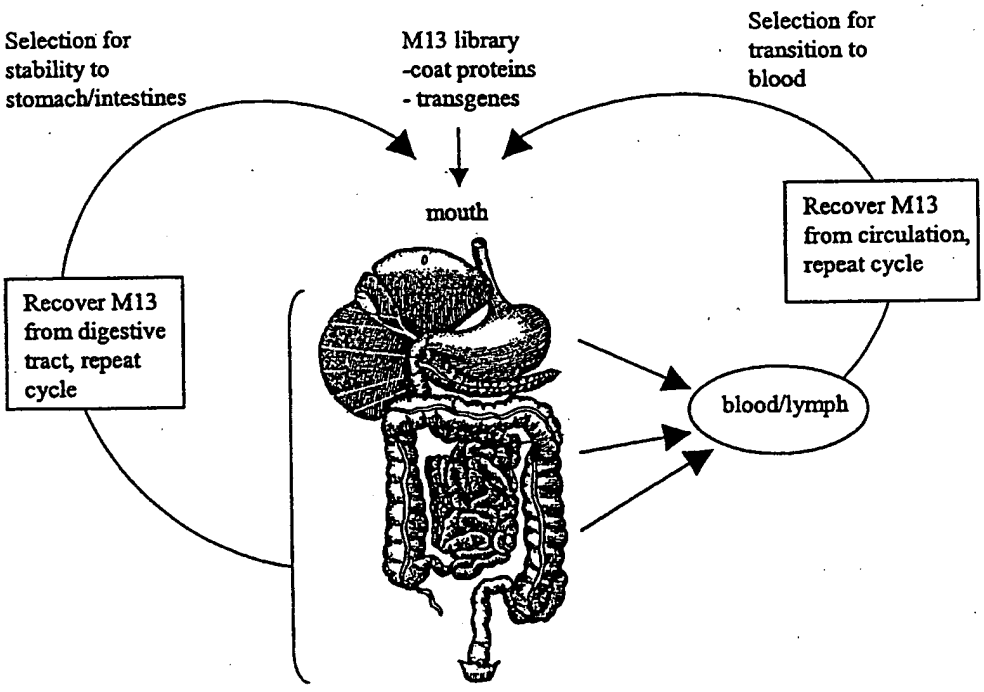


Figure 30

An alignment of the nucleotide sequences of two human CMV strains and one monkey strain.

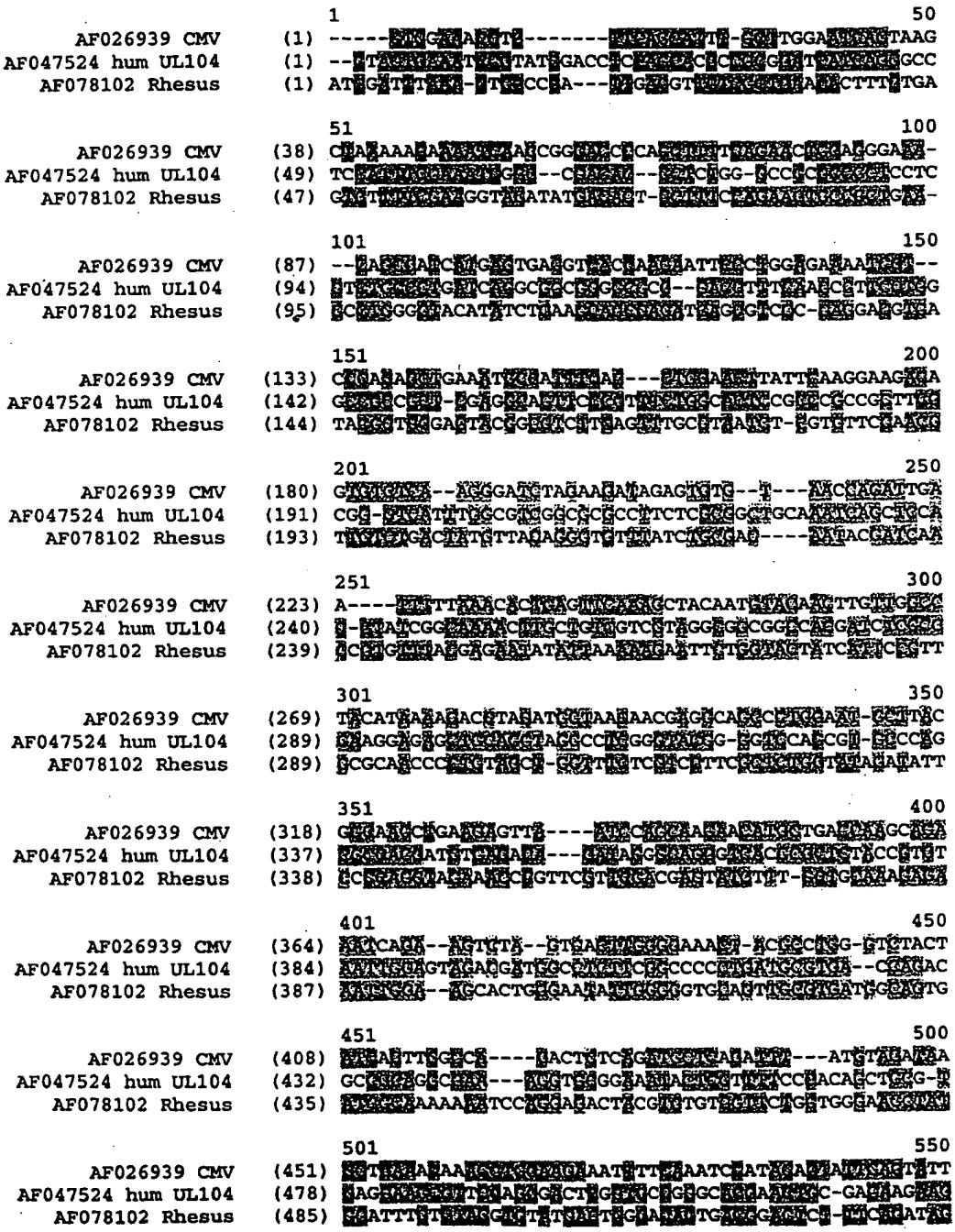


Figure 30 continued

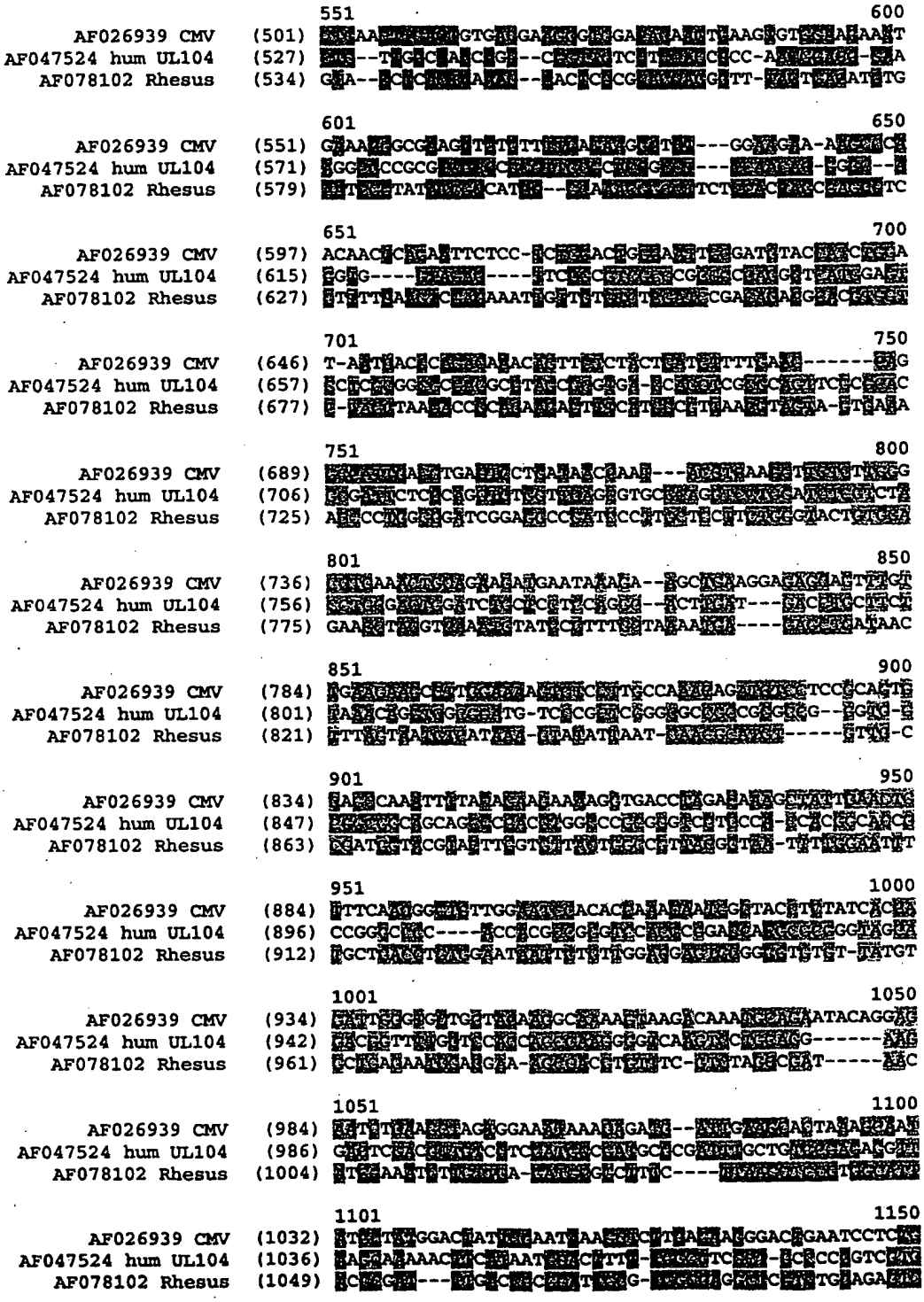


Figure 30 continued

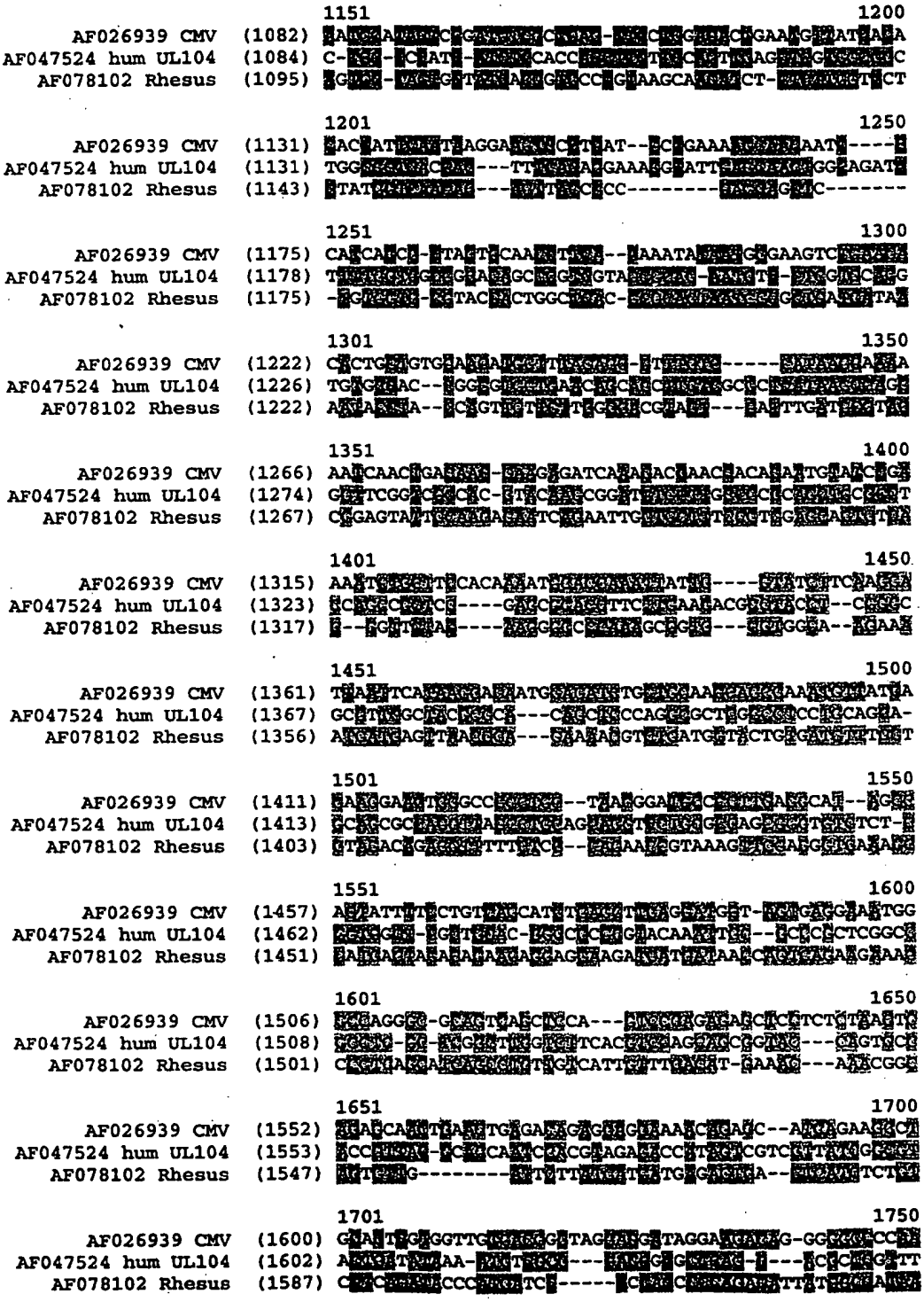


Figure 30 continued

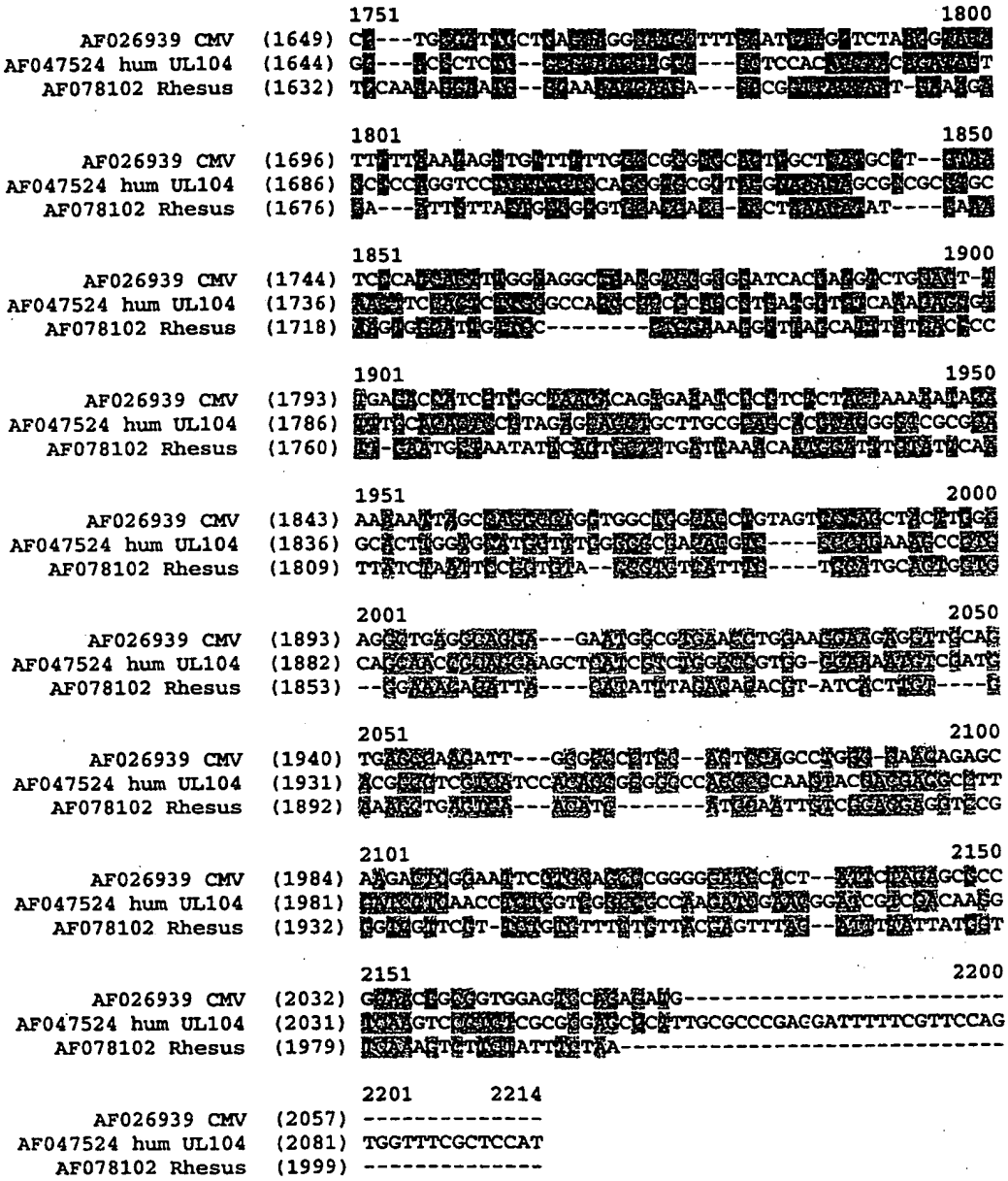
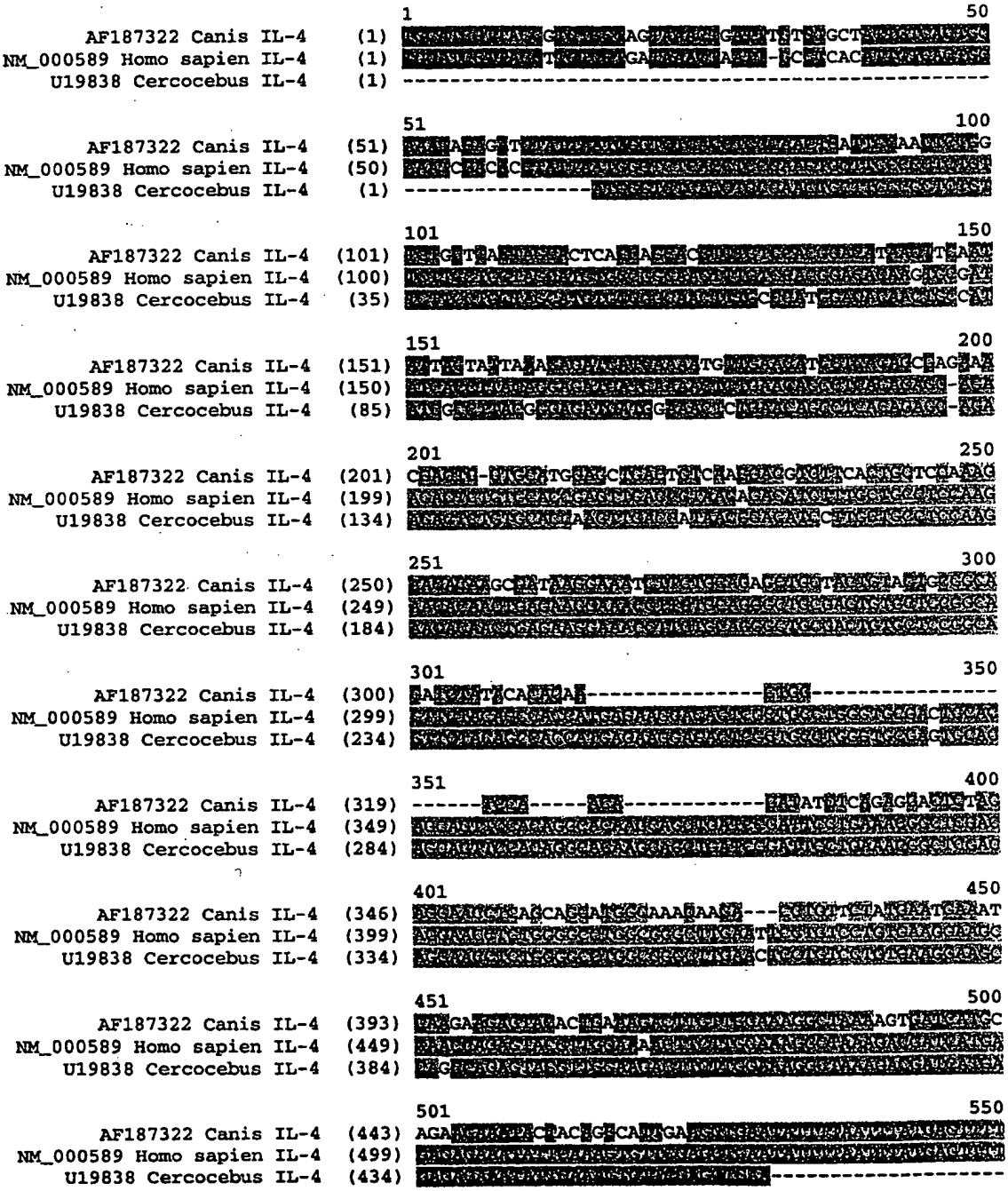


Figure 31

An alignment of IL-4 nucleotide sequences from 3 species (human, primate, and canine).



Figue 31 continued

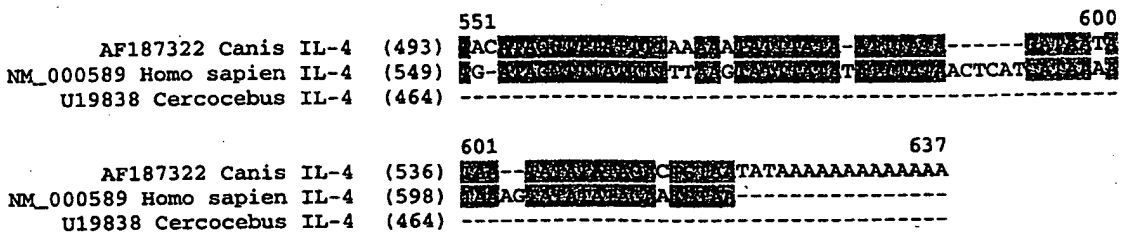


Figure 32

Evolution of polypeptides by synthesizing (in vivo or in vitro) corresponding deduced polynucleotides and subjecting the deduced polynucleotides to directed evolution and expression screening subsequently expressed polypeptides.

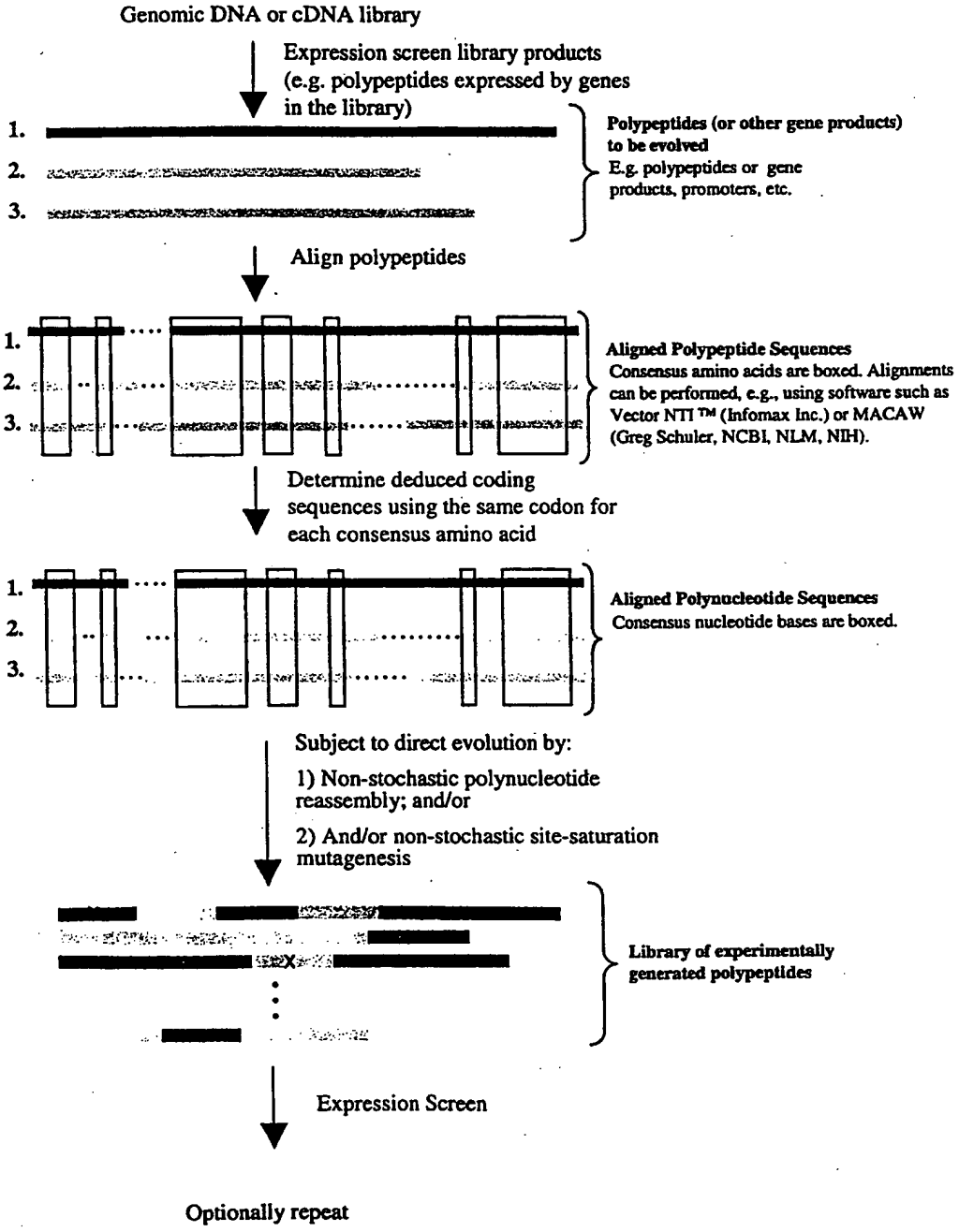


Figure 33

Directed evolution of polynucleotides (e.g. promoter sequences)

This figure shows an example of the application of non-stochastic site-saturation mutagenesis in combination with non-stochastic reassembly (e.g. oligo-directed CpG deletion(s) and/or addition(s))

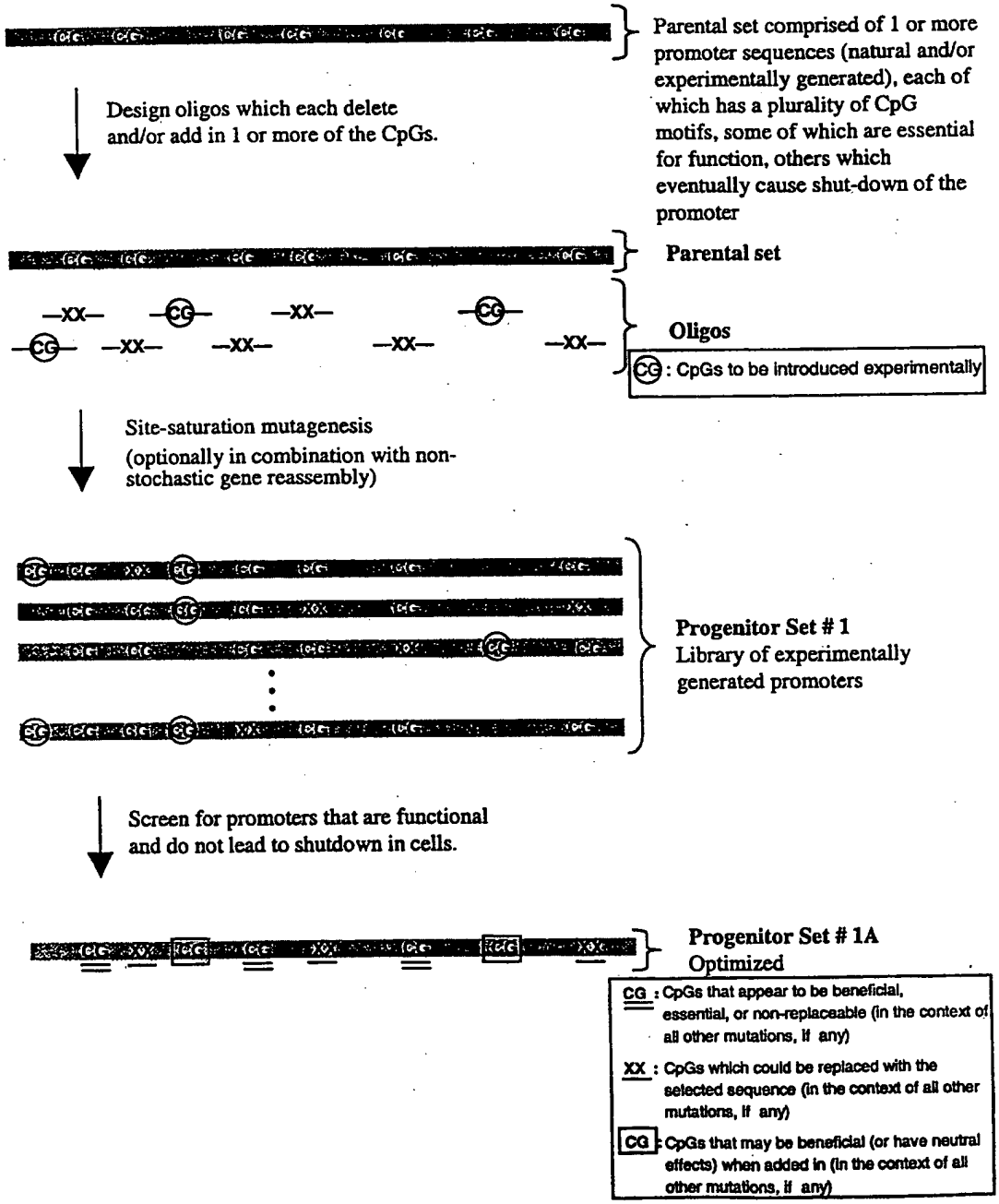


Figure 34

An example of a CTIS obtained from HbsAg polypeptide (PreS2 plus S regions).

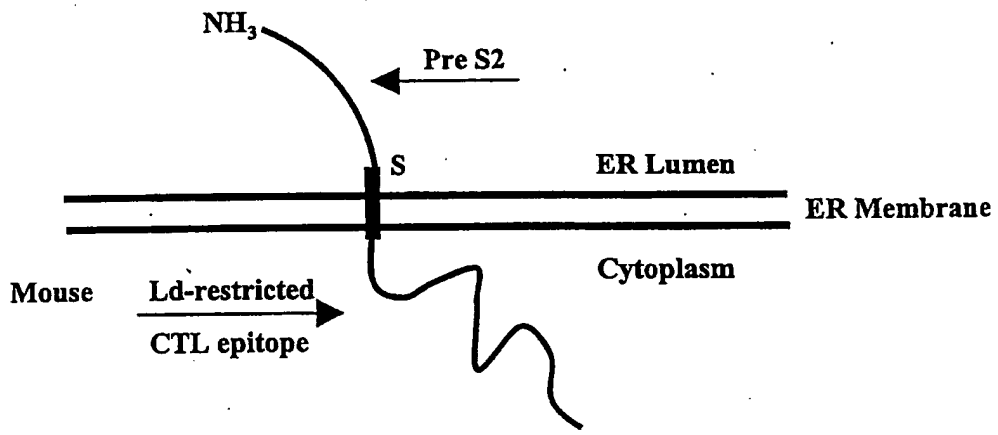


Figure 35

An example of a CTIS having heterologous epitopes attached to the cytoplasmic portion.

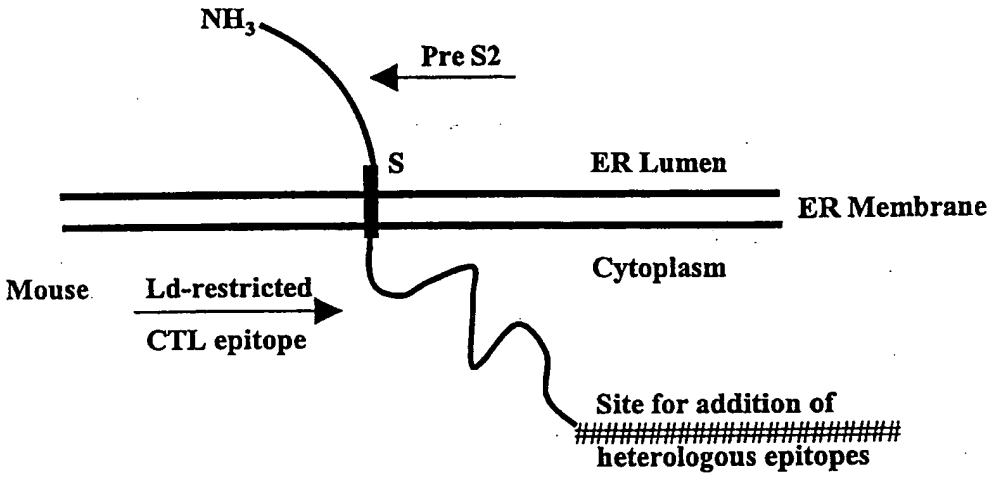


Figure 36

Method for preparing immunogenic agonist sequences (IAS).

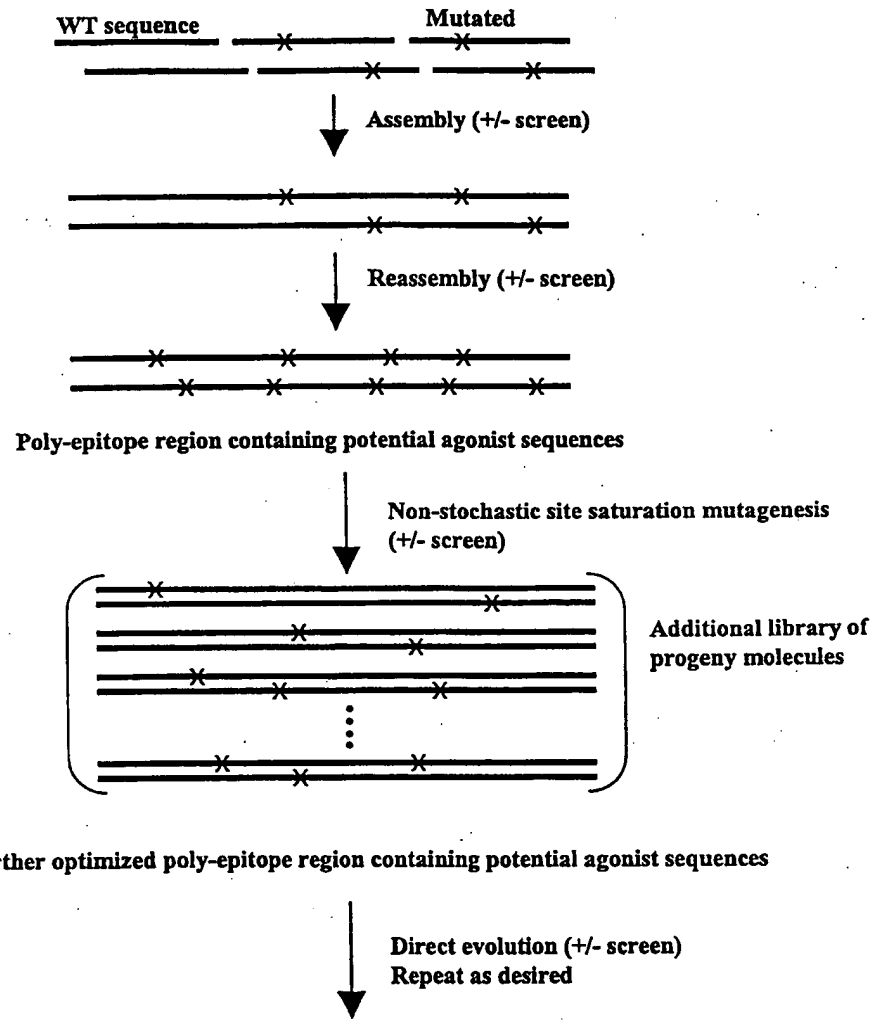


Figure 37

Improving Immunostimulatory Sequences (ISS) Using Directed Evolution.

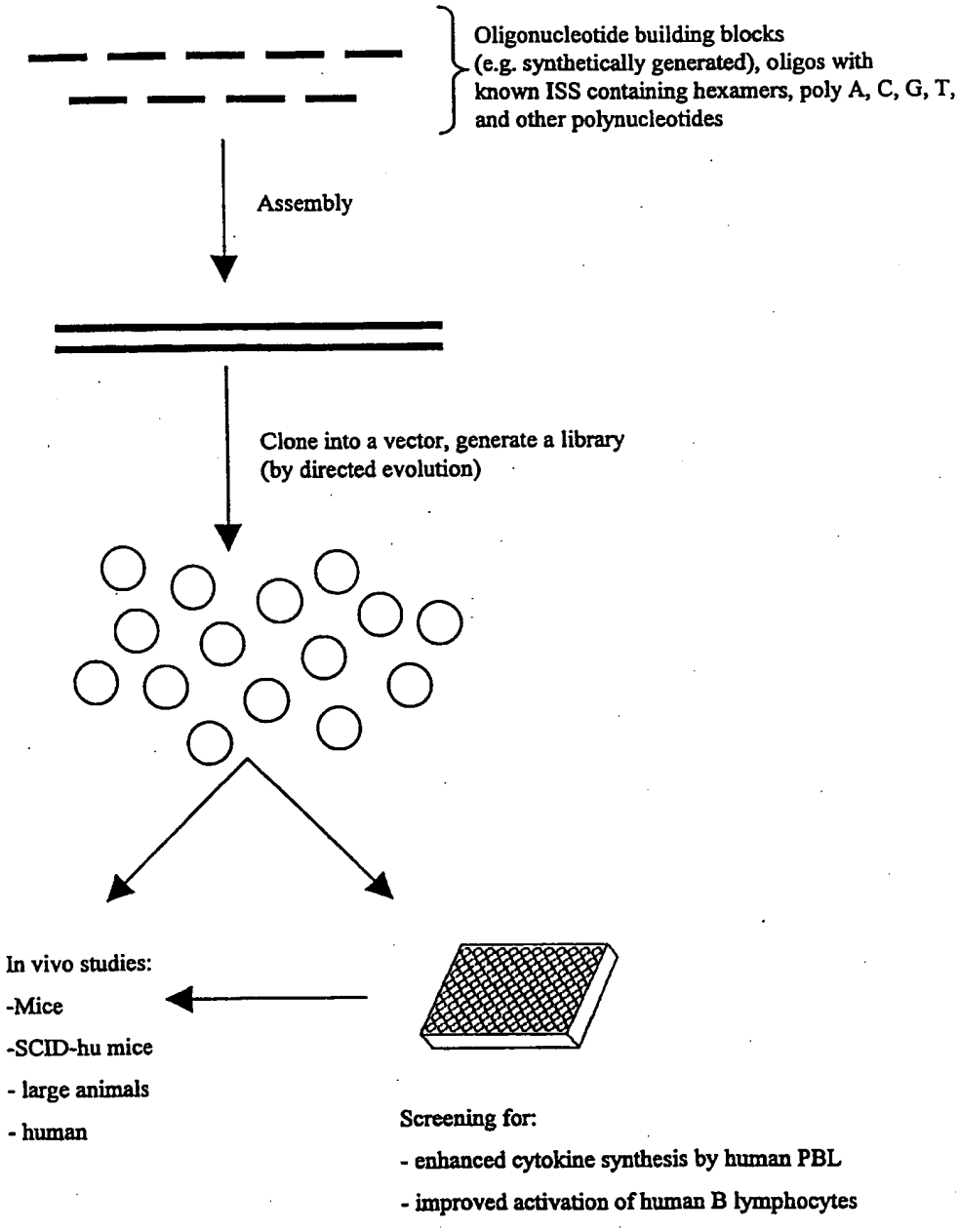


Figure 38

Screening to identify IL-12 genes that encode recombinant IL-12 having an increased ability to induce T Cell proliferation.

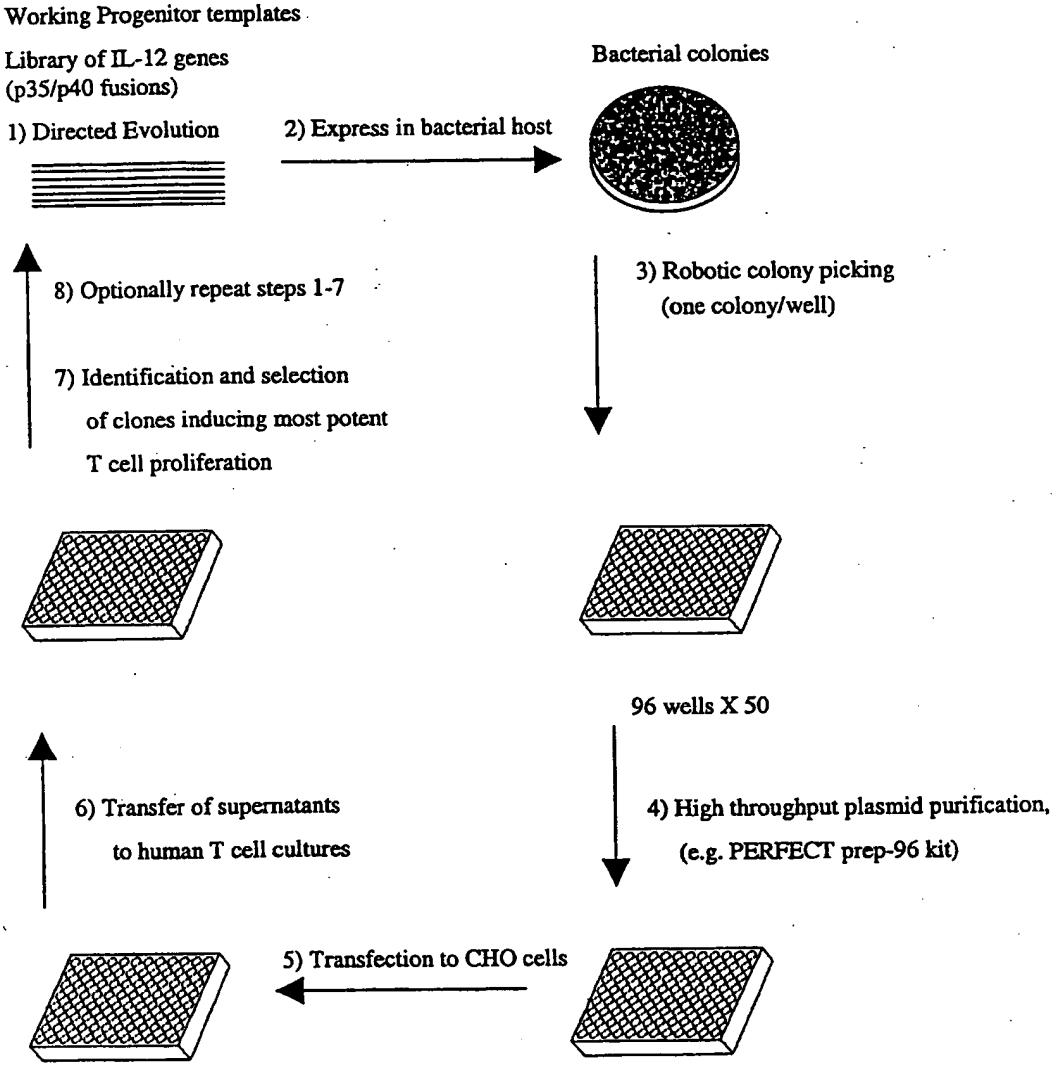


Figure 39

Model of induction of T cell activation or anergy by genetic vaccine vectors encoding different CD80 and/or CD86 variants.

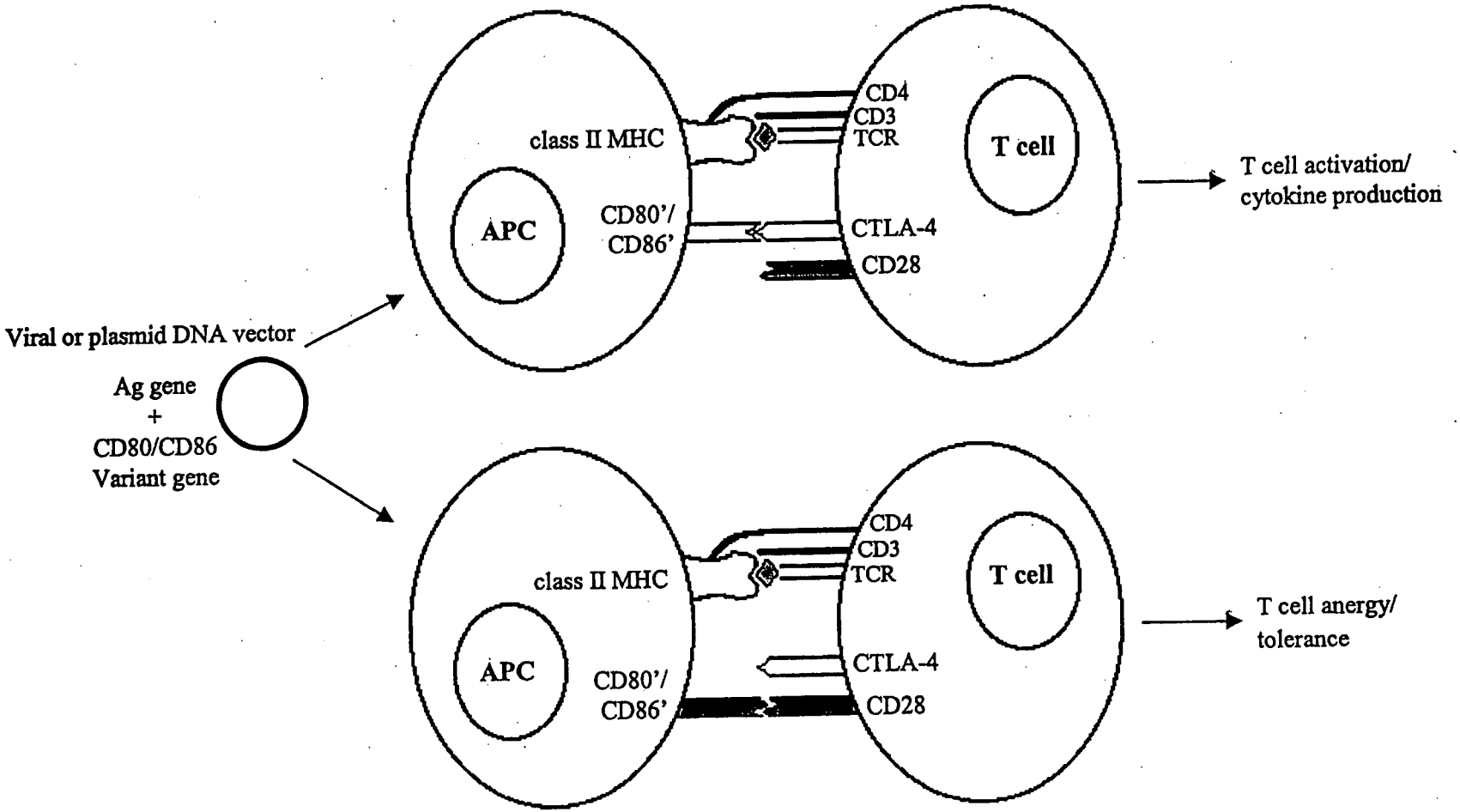


Figure 40

Screening to identify CD80/CD86 chimeric genes having an improved capacity to to induce T Cell activation or anergy.

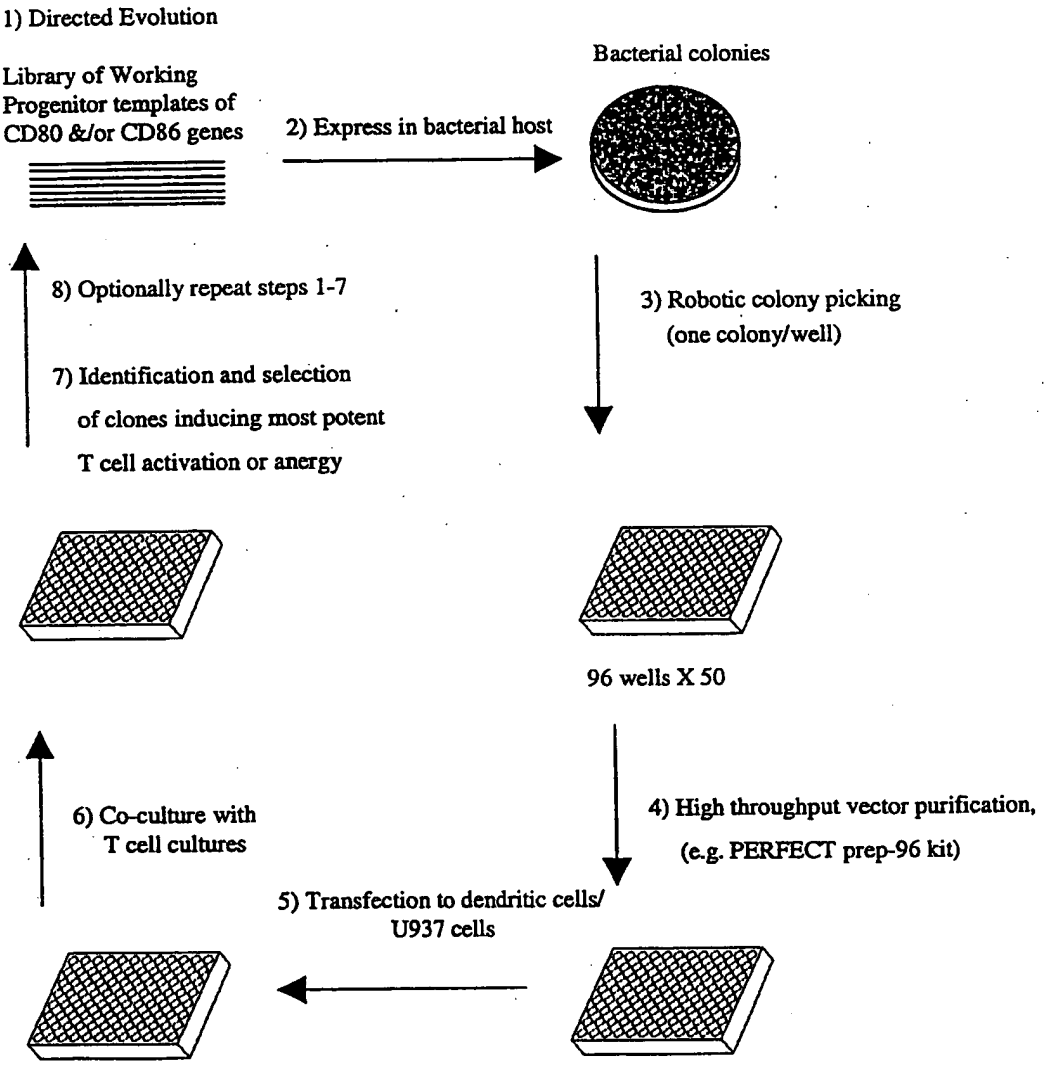


Figure 41

Figure 41. An alignment of two CMV-derived nucleotide sequences from human and primate species.

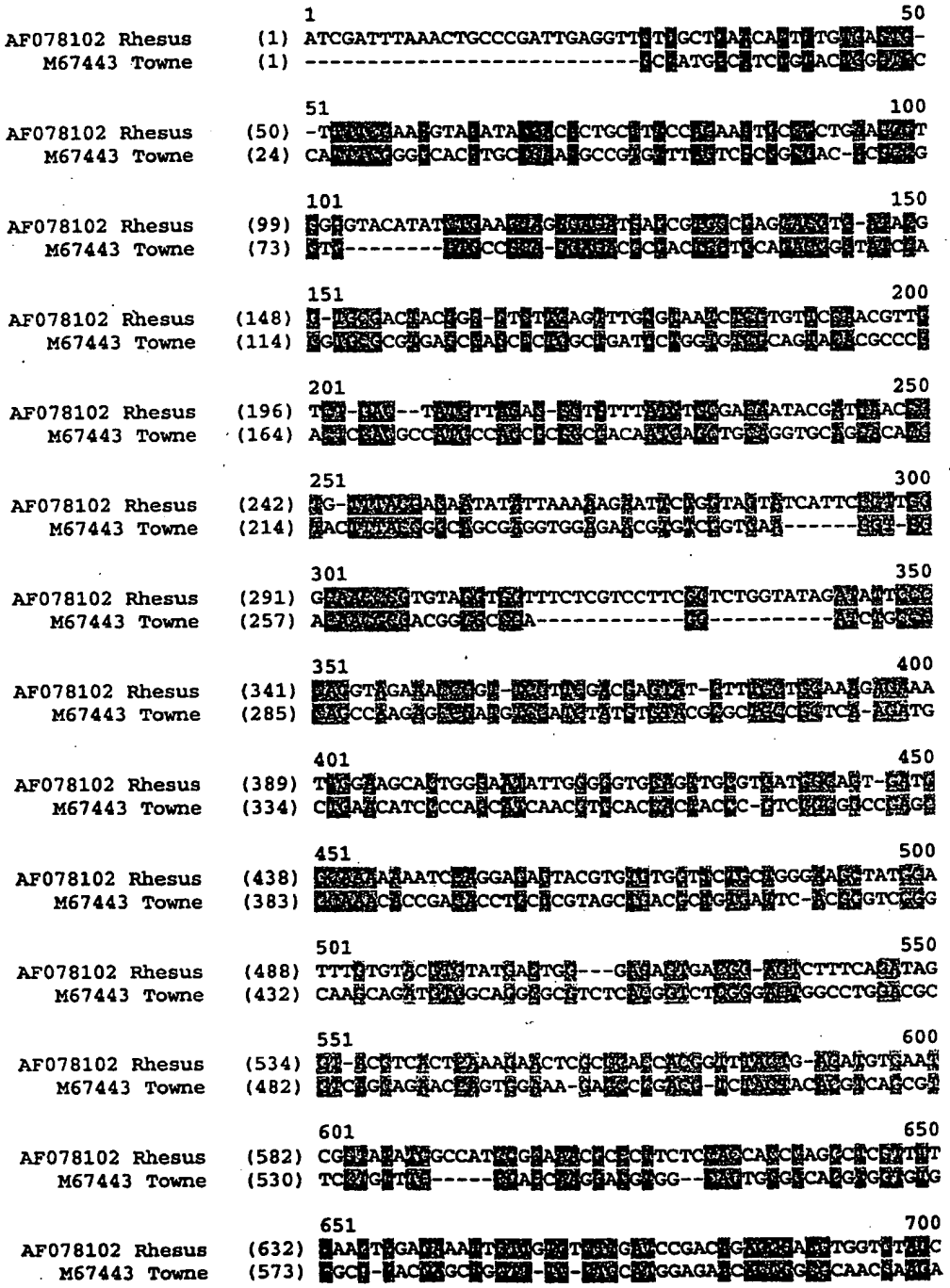


Figure 41 continued

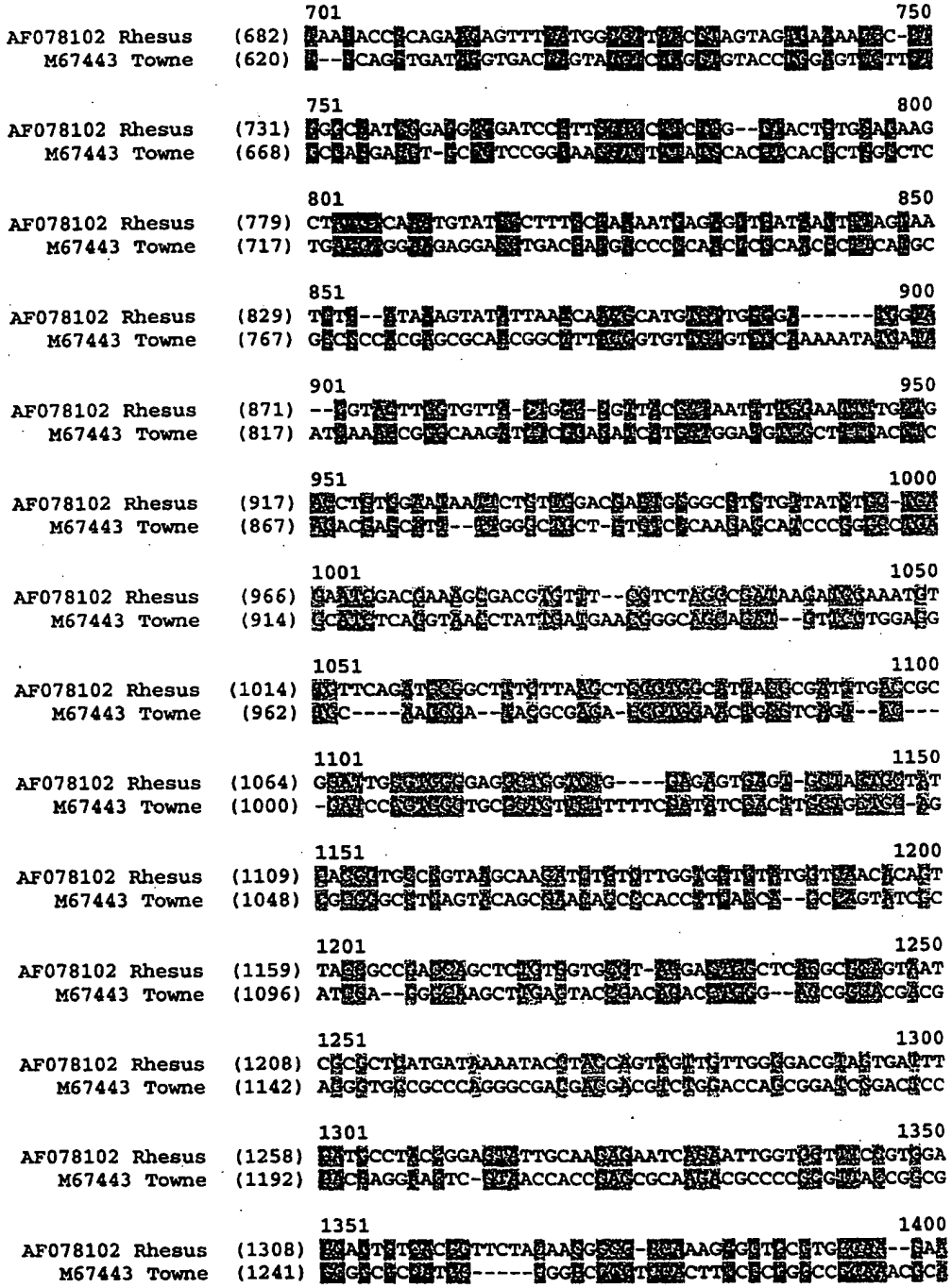


Figure 41 continued

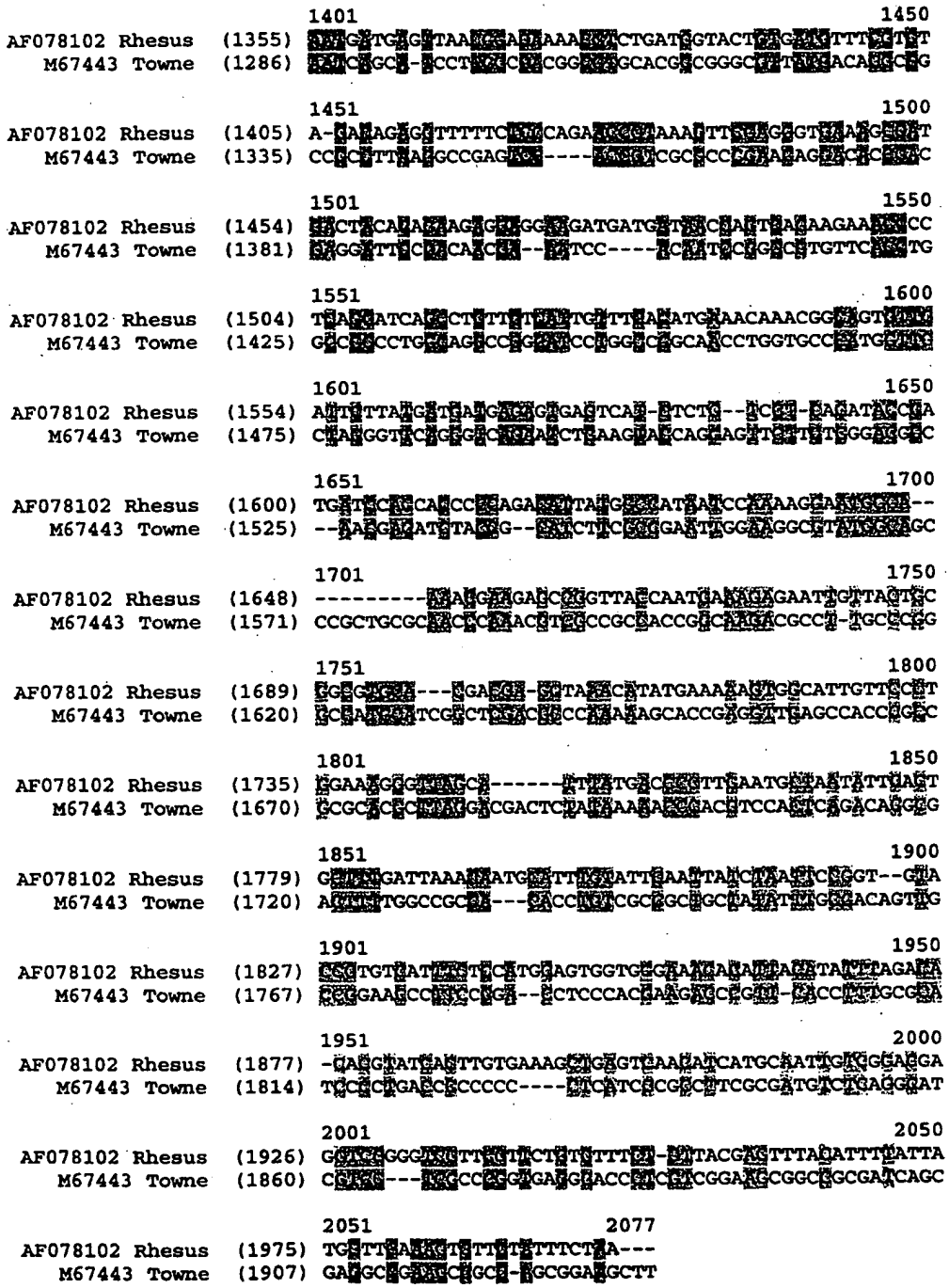


Figure 42

Figure 42: An alignment of the IFN-gamma nucleotide sequences from human, cat, rodent species.